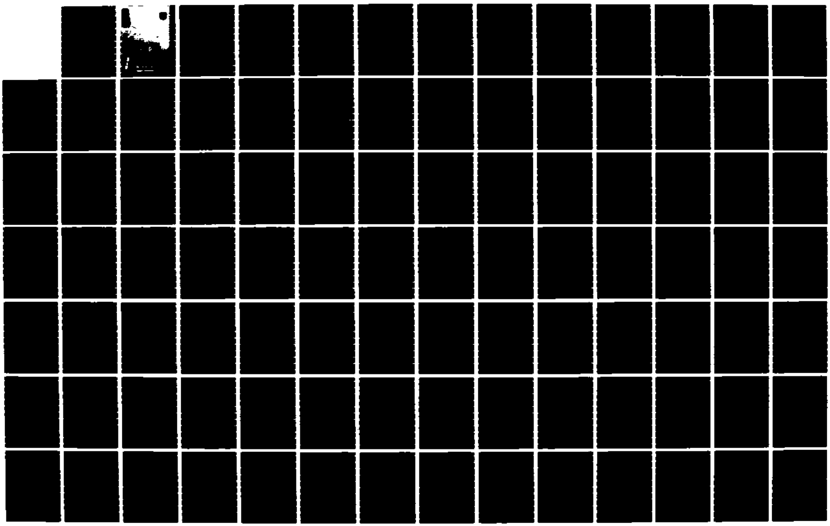
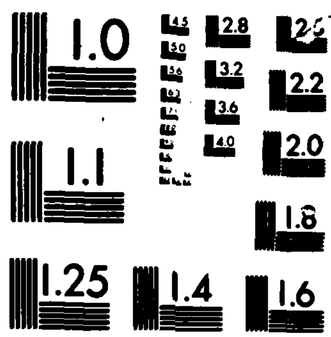
AD-A168 819 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 1/3
AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN JAN 86 SR-84(1985) N00014-83-K-0083

UNCLASSIFIED F/G 5/7 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.8
3.2
3.6
4.0
4.5
5.0
5.6

SR-84 (1985)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 October - 31 December 1985

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

(The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order number of previous Status Reports.)

ACKNOWLEDGMENTS

The research reported here was made possible
in part by support from the following sources:

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Grant HD-01994

NATIONAL INSTITUTE OF CHILD HEALTH AND HUMAN DEVELOPMENT

Contract NO1-HD-5-2910

NATIONAL INSTITUTES OF HEALTH

Biomedical Research Support Grant RR-05596

NATIONAL SCIENCE FOUNDATION

Grant BNS-8111470

NATIONAL INSTITUTE OF NEUROLOGICAL AND COMMUNICATIVE
DISORDERS AND STROKE

Grant NS 13870
Grant NS 13617
Grant NS 18010

OFFICE OF NAVAL RESEARCH

Contract N00014-83-K-0083

HASKINS LABORATORIES PERSONNEL IN SPEECH RESEARCH

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Thomas Baer
Fredericka Bell-Berti*
Catherine Best*
Geoffrey Bingham†
Gloria J. Borden*
Susan Brady*
Catherine P. Browman
Franklin S. Cooper*
Stephen Crain*
Robert Crowder*
Laurie B. Feldman*
Anne Fowler†
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Sarah Hawkins††
Satoshi Horiguchi[1]
Leonard Katz*
J. A. Scott Kelso
Andrea G. Levitt*
Alvin M. Liberman*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Richard McGowan
Kevin G. Munhall†††
Hiroshi Muta[1]
Susan Nittrouer††
Patrick W. Nye
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Douglas H. Whalen

Technical/Support

Michael Anstett
Philip Chagnon
Alice Dadourian
Vincent Gulisano
Donald Hailey
Raymond C. Huey*
Sabina D. Koroluk
Bruce Martin
Frank Merewether
Betty J. Myers
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Joy Armson
Dragana Barac
Sara Basson
Eric Bateson
Suzanne Boyce
Teresa Clifford
Andre Cooper
Margaret Dunn
Jo Estill
Carole E. Gelfer
Bruce Kay
Noriko Kobayashi
Rena A. Krakow
Deborah Kuglitsch
Hwei-Bing Lin
Katrina Lukatela
Harriet Magen
Sharon Manuel
Jerry McRoberts
Lawrence D. Rosenblum
Arlyne Russo
Richard C. Schmidt
John Scholz
Robin Seider
Suzanne Smith
Katyanee Svastikula
David Williams

---

*Part-time
[1]Visiting from University of Tokyo, Japan
†NIH Research Fellow
††NRSA Training Fellow
†††Natural Sciences and Engineering Research Council of Canada Fellow

CONTENTS

Status Report on Speech Research

Haskins Laboratories

TASK DYNAMIC COORDINATION OF THE SPEECH ARTICULATORS: A PRELIMINARY MODEL*

Elliot Saltzman

Abstract. A task dynamic model of skilled movements originally formulated with reference to limb tasks (Saltzman & Kelso, 1983a/in press) is extended to incorporate speech production. In the model, qualitative differences among tasks are captured by corresponding topological differences in the dynamical structures of abstract task-space control regimes. These task-dynamic regimes remain invariant throughout a given limb or speech gesture. Major levels of dynamical representation and associated coordinate transformations among these levels are introduced in a discussion of a planar reaching task for a 3-joint limb. Extensions to speech production focus on bilabial movements during tasks involving discrete closing and repetitive cyclic gestures. The discrete task shows how the model exhibits utterance-specific immediate compensation to jaw perturbations; the cyclic task shows how continuous articulator trajectories may be generated that are useful for speech synthesis. Significantly, the task-dynamic model generates coordinated articulatory movements from the simple specification of abstract dynamic parameters, and requires neither explicit trajectory planning for unperturbed movements nor explicit error detection and replanning for perturbed movements.

It is perhaps a truism that skilled actions of the limbs and speech articulators are goal directed. It is equally true, however, that such actions are performed by effector systems that are indifferent to the goals of would-be performers. An effector system is the set of limb segments or speech articulators used in a given action; a terminal device or end-effector is the part of a controlled effector system that is directly related to the goal of a performed action. Thus, in a reaching task, the fingers define the terminal device and the arm and hand comprise the effector system; in a "cup-to-mouth" task, the grasped cup is the terminal device and the combination of hand and arm constitutes the effector system; in a steady-state vowel production task, the tongue body is the terminal device and the jaw and tongue comprise the effector system. During skilled actions, the numerous degrees of freedom defined by the muscles and joints of such effector systems must be harnessed functionally in a manner specific to the task or goal at hand.

In addition to a skill's goal directedness, it is also clear that ordinary actions (such as walking or talking) or extraordinary actions (such as ballet or operatic singing) are never performed twice in exactly the same way. Yet observers and students of such activities seem to share the

---

*Experimental Brain Research Supplementum, in press.

Acknowledgment. This research was supported in part by NIH Grant NS-13617 and Contract No. N0014-83-K-0083 from the U. S. Office of Naval Research.

intuition that there is a task-specific commonality or invariance that underlies the separate task performances. In the present paper, a theoretical approach to these dual issues of contextual variation and task-specific invariance in skilled actions is described. This approach is called task dynamics (Saltzman & Kelso, 1983a/in press), and promises to provide within a single framework a parsimonious account of both variable and invariant aspects of well-learned, skilled actions. Saltzman and Kelso (1983a/in press) describe how a mathematical, task-dynamic model can be applied to tasks involving relatively simple arm movements in the horizontal or sagittal planes (e.g., reaching discretely and cyclically, transporting cup-to-mouth and crank-turning). The present paper describes how task-dynamic modeling is being extended by this author and his colleagues at Haskins Laboratories to the coordination and regulation of the speech articulators during linguistically meaningful tasks (cf. Browman & Goldstein, 1985; Browman, Goldstein, Kelso, Rubin, & Saltzman, 1984; Kelso, Vatikiotis-Bateson, Saltzman, & Kay, 1985).

There are (at least) two signature properties of skilled actions--trajectory shaping and immediate compensation--that must be accounted for by a theory of coordination and control. Trajectory shaping refers to the tendency of end effector trajectories to display forms that are characteristic of the demands of performed tasks. For example, it has been demonstrated in several laboratories that in planar reaching tasks using the shoulder and elbow joints, the hand moves in a quasi-straight line toward the target (e.g., Bizzi & Abend, 1982; Bizzi, Accornero, Chapple, & Hogan, 1981; Morasso, 1981; Soechting & Lacquaniti 1981; Wadman, Denier van der Gon, & Derkson, 1980; see also Hollerbach & Atkeson, in press). Similarly, in cup-to-mouth tasks, the grasped cup must maintain a spillage-preventing horizontal orientation while en route from table to mouth.

The second characteristic of skilled gestures, immediate compensation, refers to the task-specific flexibility of action systems in reorganizing themselves when faced with unexpected disturbances or perturbations. Thus, compensation for the perturbation of a given effector during a movement trajectory is achieved by readjusting the activity over the entire system in order to achieve the task goal (e.g., Bernstein, 1967; Marsden, Merton, & Morton, 1983; Nashner & McCollum, 1985). Further, these readjustments appear to occur automatically without the need to detect the disturbance explicitly, replan a new movement, and execute the new movement plan. Kelso, Tuller, V.-Bateson, and Fowler (1984) have demonstrated such behavior in the speech articulators (jaw, upper and lower lip, tongue body) when subjects produced the utterances /bæb/ or /bæz/ across a series of trials in which the jaw was occasionally and unpredictably tugged downward while moving upward to the final /b/ or /z/ constriction (see also Abbs & Gracco, 1983; Folkins & Abbs, 1975). The system's response to the jaw perturbation was measured by observing the motions of the jaw and upper and lower lips as well as the electromyographic (EMG) activities of the orbicularis oris superior (upper lip), orbicularis oris inferior (lower lip), and genioglossus (tongue body) muscles. The investigators found relatively "immediate" task-specific compensation (i.e., 20-30 ms from onset of jaw pull to onset of compensatory response) in remote articulators to jaw perturbation. For /bæb/ (in which final lip closure is crucial) they found increased upper lip activity (motion and EMG) relative to the unperturbed control trials but normal tongue activity; for /bæz/ (in which final tongue-palate constriction is important) they found increased tongue activity relative to controls, but normal upper

lip motion. The speed of these task-specific patterns suggests that compensation does not occur according to traditionally defined "intentional" reaction time processes, but rather according to an automatic, "reflexive" type of organization. However, such an organization is not defined in a hard-wired input/output manner. Instead, these data imply the existence of a selective pattern of coupling or gating among the component articulators that is specific to the utterance produced. Such compensatory behavior represents the classic phenomenon of motor equivalence (Hebb, 1949; Lashley, 1930) according to which a system will find alternate routes to a given goal if an initially intended route is unexpectedly blocked.

What type of coordinative processes could generate, in a task-specific manner, both characteristic trajectory patterns for unperturbed movements and spontaneous, compensatory behaviors for perturbed movements? The task-dynamic model for effector systems having many articulatory degrees of freedom was developed in an effort to deal with these issues (Saltzman & Kelso, 1983a/in press; see also Boylls & Greene, 1984, for related discussions of task-specific dynamics). The model is labeled task dynamic since: a) it deals with the performance of well-learned skilled movements or gestures designed to accomplish real-world tasks; and b) it is defined with respect to the dynamics that underlie a given action's kinematics. Note that kinematics refers to a gesture's observable spatiotemporal properties (e.g., its position, velocity, and acceleration trajectories over time), while dynamics refers to the pattern of the underlying field of forces that gives rise to these kinematics. The task-dynamic approach extends and elaborates the view that the functional units of action (or <u>coordinative structures</u>; e.g., Easton, 1972; Fowler, 1977; Kelso, Southard, & Goodman, 1979; Turvey, 1977) underlying the performance of a given gesture may be identified with abstractly defined, task-specific control regimes whose dynamic parameters (e.g., stiffness, damping, rest position) remain constant over the course of the gesture (cf. Fitch & Turvey, 1978; Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980, 1982; Saltzman & Kelso, 1985). In the task-dynamic model, the control regime that governs the performance of a particular gesture or task is defined functionally as an abstract (<u>task space</u>) dynamical system that is effector-independent, i.e., it does <u>not</u> explicitly incorporate the particular end-effectors directly involved in performing the task. It is hypothesized that a common task-space description underlies the functional equivalence of different effector systems for the performance of a given task, e.g., writing one's signature using a pencil held in the hand or between the teeth. Relatedly, qualitative differences between tasks are captured by corresponding <u>topological</u> distinctions among task-space dynamical systems (see also Arbib, 1984, for a related discussion of the relation between task and controller structures).

For example, gestures involving a hand's discrete motion to a single spatial target and repetitive cyclic motion between two such targets are characterized by <u>point attractor</u> and <u>periodic attractor</u> dynamical regimes, respectively (cf. Abraham & Shaw, 1982). The behaviors of these two types of dynamical systems may be represented in the phase plane (i.e., where system velocity is plotted vs. position) as illustrated in Figure 1, along with examples of corresponding equations of motion. Figure 1A shows a point attractor regime characterized by an (underdamped) mass-spring equation of motion. This system displays <u>point stability</u> or <u>equifinality</u>, in that it will asymptotically attain the equilibrium position, x_0, regardless of initial conditions for x and $\dot{x}$ and despite any transient perturbations encountered

during its motion trajectory. Figure 1B shows a periodic attractor regime with a stable cyclic orbit (i.e., limit cycle) that is approached asymptotically by all trajectories (except those starting exactly at x_0) regardless of transiently introduced perturbations. The value of specifying a system's behavior in terms of topologically defined attractors is that such attractors provide task-specific, low dimensional descriptions for movement systems with many degrees of freedom, and promise to provide an elegant notational scheme for capturing the dynamical invariance across different effector systems that are observed to perform identical tasks. Distinct topologies correspond, therefore, to distinct patterns of task-dynamic parameters (e.g., damping and stiffness coefficients), and have been labeled the organizational invariants for skilled actions of different types (Fowler & Turvey, 1978; Saltzman & Kelso, 1983a/in press, 1983b). Such patterns denote functions that are preserved invariantly over changes in the parameters' specific values. In the task-dynamic model, the values of these tuning parameters (e.g., Greene, 1972; Saltzman & Kelso, 1983a/in press, 1983b) are determined according to factors such as the rate or amplitude of movement, and are defined to be constant over the course of a given gesture.

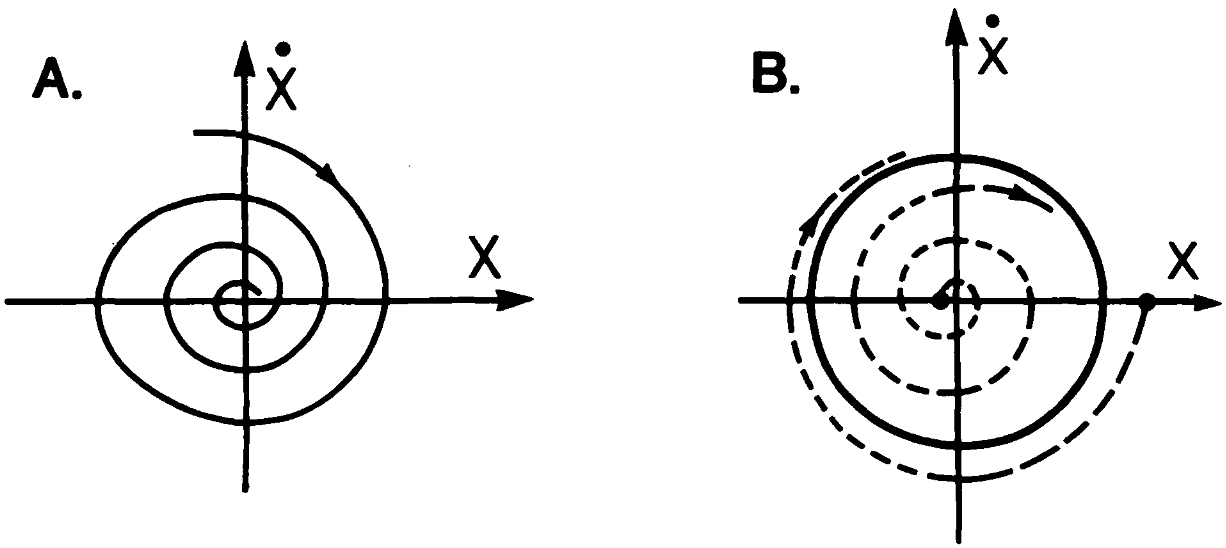


Figure 1. Representative phase plane $(x,\dot{x})$ trajectories for point attractor (A) and periodic attractor (B) systems. Examples of motion equations are: A. $m\ddot{x} + b\dot{x} + kx = 0$, where m = mass, b = damping, k = stiffness; B. $m\ddot{x} + b\dot{x} + kx = f(x,\dot{x})$, where $f(x,\dot{x})$ is a nonlinear damping (i.e., escapement) term, and all other coefficients are as in A.

The task-dynamic model is able to account for the phenomena of trajectory shaping and immediate compensation without the need for explicit trajectory planning or replanning (see Saltzman & Kelso, 1983a/in press, for further details). Note that defining invariant patterns of dynamic parameters at the level of articulatory degrees of freedom (e.g., stiffness and damping parameters at the joints of an arm) will not suffice to generate these behaviors. Constant articulatory-dynamic parameters will not generate the quasi-straight-line hand trajectories seen in planar reaching tasks (Delatizky, 1982; Hollerbach, 1982); rather, such trajectory shapes must result from task-specific patterns of change in these parameters during the reaching gestures. Similarly, the immediate compensation data for speech described above (Kelso et al., 1984) could not be generated by a system with a constant rest configuration parameter (i.e., a vector whose components are constant rest positions for the lips and jaw). As shown in these data, when sustained perturbations were introduced during articulatory closing gestures, the system "automatically" achieved the same constriction goals as for unperturbed gestures, but with different final or rest configurations. Thus, both trajectory shaping and immediate compensation behaviors appear to result from the way that dynamic parameters at the articulatory level are constrained to change during a gesture in a context-dependent manner. In the task-dynamic model, such patterns of constraint originate in corresponding invariant patterns of dynamic parameters at the task-space level of description.

Example 1: Planar Reaching, 3 joints. Using, for illustrative purposes, a discrete reaching task in the horizontal plane with angular motion at the shoulder, elbow, and wrist joints, the operation of a given task-dynamic regime may be understood in the following way. First, the functional aspects of a reaching gesture are specified in a two-dimensional task space as an invariant point attractor (e.g., a two-dimensional damped mass-spring system; see Figure 2A). These dynamics give rise to an evolving pattern of state-dependent "forces" exerted on an effector-independent terminal device (i.e., a task mass). In the task space, the reach target defines the origin of a Cartesian coordinate system, with axis t_1 ("Reach" axis) defined along a line from the initial position of the task mass to the target, and axis t_2 ("Normal" axis) defined normal to t_1. The equations of motion for this task-dynamic regime are described in matrix notation as follows:

$$M_T\ddot{\underset{\sim}{t}} + B_T\dot{\underset{\sim}{t}} + K_T\underset{\sim}{t} = 0 \text{ , where} \tag{1}$$

$$M_T = \begin{bmatrix} m_T & 0 \\ 0 & m_T \end{bmatrix}; \quad B_T = \begin{bmatrix} b_{T1} & 0 \\ 0 & b_{T2} \end{bmatrix};$$

$$K_T = \begin{bmatrix} k_{T1} & 0 \\ 0 & k_{T2} \end{bmatrix};$$

m_T = task-mass coefficient;

b_{T1}, b_{T2} = damping coefficients;

k_{T1}, k_{T2} = stiffness coefficients.

Equation (1) describes a linear, uncoupled set of task-space equations, whose terms are defined in units of force, and whose dynamic parameters (i.e., M_T, B_T, K_T) are constant. In Figure 2A the corresponding damping and

stiffness elements are represented in lumped form by the squiggles in the lines connecting the task mass to axes t_1 and t_2.

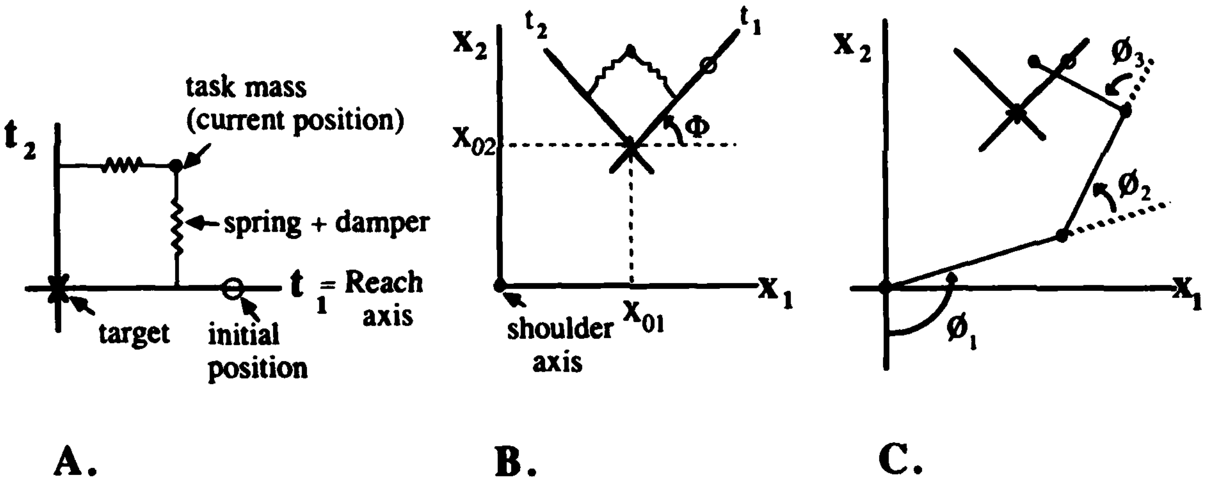


Figure 2. Discrete reaching: A. Task space ($\underset{\sim}{t}$); B. Shoulder space ($\underset{\sim}{x}$). Task space is located and oriented in shoulder-centered reference frame via $\underset{\sim}{x}_0$ and Φ, respectively; C. Model articulator space ($\underset{\sim}{\phi}$). ϕ's denote joint angles.

Second, the task mass is identified with the relevant "virtual" end-effector (e.g., a virtual finger tip), and the task-space dynamic system is transformed kinematically into a two-dimensional body-space system (x_1, x_2; shoulder space) governing movements of the virtual end-effector (see Figure 2B). Thus, the task space is located and oriented in body-space coordinates according to the tuning parameters $\underset{\sim}{x}_0$ (the body-space position vector of the task-space origin) and Φ (the orientation angle between task axis t_1 and body axis x_1), respectively. The resulting set of linear body-space equations of motion for the task's terminal device are defined in matrix form as follows (Note: In these and the following equations, a superscript T denotes the vector or matrix transpose operation):

$$M_B\ddot{\underset{\sim}{x}} + B_B\dot{\underset{\sim}{x}} + K_B\Delta\underset{\sim}{x} = \underset{\sim}{0}, \text{ where} \tag{2}$$

$M_B = M_T R$, where M_T = task-space mass matrix; and

R = the rotation transformation matrix with elements $r_{ij}(\Phi)$ converting task-space variables into body-space form;

$B_B = B_T R$, where B_T = task-space damping matrix;

$K_B = K_TR$, where K_T = task-space stiffness matrix; and

$\Delta x = x - x_0$, where $x = (x_1, x_2)^T$, the current body-space position vector of the terminal device.

Note that Equation (2), unlike Equation (1), represents a set of body-space equations that are (usually) coupled due to the rotation transformation (i.e., the off-diagonal matrix elements are generally non-zero). However, as with the task-space equations, the terms of (2) are defined in force units and the resultant set of body-space dynamic parameters is constant.

Third, the body-space dynamic system is transformed into a three dimensional "model" articulator space where the moving segments (upper arm, forearm, and hand) have lengths but are massless (see Figure 2C). Like the transformation from task space to body space, this transformation is a strictly kinematic one (since the segments have no mass) and involves only the substitution of variables defined in one coordinate system for variables defined in another coordinate system. As illustrated in Figure 2C, this corresponds to expressing body-space variables (x, $\dot{x}$, $\ddot{x}$) as functions of an arm model's kinematic variables (ϕ, $\dot{\phi}$, $\ddot{\phi}$; where $\phi = [\phi_1, \phi_2, \phi_3]^T$, and ϕ_1 = shoulder angle defined relative to axis x_2, ϕ_2 = elbow angle defined relative to the upper arm segment, ϕ_3 = wrist angle defined relative to the forearm segment), and the arm's proximal (shoulder) and distal (finger tip) ends are attached to the body space origin and the terminal device/task mass, respectively. The body-space variables of Equation (2) are transformed into the joint-angle variables of the massless arm model using the following kinematic relationships:

$$x = x(\phi) \tag{3a}$$

$$\dot{x} = J(\phi)\dot{\phi} \tag{3b}$$

$$\ddot{x} = J(\phi)\ddot{\phi} + (dJ(\phi)/dt)\dot{\phi} \tag{3c}$$

$$= J(\phi)\ddot{\phi} + V(\phi)\dot{\phi}_p, \text{ where}$$

$x(\phi) = (x_1(\phi), x_2(\phi))^T$, the current body-space position vector of the terminal device expressed as a function of the current model arm configuration;

$\dot{\phi}_p = [\dot{\phi}_1^2, \dot{\phi}_1\dot{\phi}_2, \dot{\phi}_1\dot{\phi}_3, \dot{\phi}_2^2, \dot{\phi}_2\dot{\phi}_3, \dot{\phi}_3^2]^T$ the current model arm joint velocity product vector;

$J(\phi)$ = the <u>Jacobian</u> transformation matrix whose elements J_{ij} are partial derivatives, $\partial x_i/\partial\phi_j$, evaluated at the current ϕ; and

$V(\phi)$ = a matrix resulting from rearranging the terms of the expression $(dJ(\phi)/dt)\dot{\phi}$ in order to segregate the joint velocity products into a single vector $\dot{\phi}_p$.

Using the kinematic relationships in Equation 3, the model effector system's equation of motion is as follows:

$$M_BJ\ddot{\phi} + B_BJ\dot{\phi} + K_B\Delta x(\phi) = -M_BV\dot{\phi}_p, \text{ where} \tag{4}$$

M_B, B_B, K_B are the same constant matrices used in Equation (2); and

$\Delta x(\phi) = x(\phi) - x_0$, where x_0 = the same constant vector used in Equation (2); It should be noted that since Δx in Equations 2 and 4 is not assumed to be "small," a differential approximation $dx = J(\phi)d\phi$ is not justified and, therefore, Equation (3a) was used instead for the kinematic displacement transformation into model arm variables.

The terms of (4) are still defined in units of force, not torque, and may be rewritten in units of angular acceleration:

$$\ddot{\phi} + J^* M_B^{-1} B_B J \phi + J^* M_B^{-1} K_B \Delta x(\phi) + J^* V \phi_p = 0, \quad \text{where} \qquad (5)$$

J^* is a weighted Jacobian pseudoinverse (e.g., Benati, Gaglio, Morasso, Tagliasco, & Zaccaria, 1980; Klein & Huang, 1983; Whitney, 1972) that is used because there are a greater number of model articulator variables than spatial variables for this task. Hence, the model effector system is redundant (e.g., Saltzman, 1979), the inverse kinematic transform from spatial to model articulator coordinates is indeterminate, and the Jacobian inverse (J^{-1}) cannot be defined. More specifically, $J^* = A^{-1}J^T(JA^{-1}J^T)^{-1}$, where A is a positive definite articulatory weighting matrix whose elements are constant during a given gesture. Using J^* provides a unique, optimal least squares solution for the differential transformation from body-space to model articulator variables that is weighted according to the pattern of elements in the A-matrix. In current modeling, the A-matrix is defined to be of diagonal form, and a given set of articulator weights will constrain motion of an articulator in direct proportion to the magnitude of the corresponding weighting element. Hence, different articulator weighting patterns are associated with different patterns of relative angular motions of the three joints for the same task- space motion of the task mass (or body-space motion of the virtual fingertip). For example, one weighting pattern might correspond to predominant shoulder motion, while a second weighting pattern might correspond to predominant elbow motion for the same task- or body-space trajectory of the terminal device. In this sense, elements of the A-matrices used in the associated J^*'s define a further set of tuning parameters for the model effector system's equation of motion (Equation 5).

The task-dynamic model allows one to define for the discrete reach (as well as other tasks) an invariant task-space dynamic regime that: a) is specified by a constant set of task-dynamic parameters; and b) constrains in a context-dependent way the evolving pattern of changes in the model arm's articulatory-dynamic parameters (i.e., stiffnesses, damping and equilibrium positions of shoulder, elbow and wrist joints) during the course of the gesture. Thus, one may interpret the task-specific, coherent movements of the model effector system as resulting from the way that instantaneous task-space "forces" acting on the associated terminal device are distributed across the model arm's articulatory degrees of freedom during the course of the planar reach. At any given instant during this gesture, the partitioning is based on two factors:

a) the task-specific, constant set of task space (Equation 1), body space (Equation 2), and model articulator space (Equations 4 and 5) dynamic parameters; and

b) the current values of elements in the posturally dependent transformation matrices (i.e., the J and J^T matrices in Equations 4 and 5) that relate motions of the articulators at their current configuration to corresponding body-space motions of the virtual fingertip. Because these elements are nonlinear functions of the current arm model posture, $\boldsymbol{\phi}$, the elements of the matrix products in Equations 4 and 5 (i.e., the coefficients that define articulatory-dynamic parameters) are also dependent on the evolving configuration of the arm model.

The final step in the task-dynamic approach is to exploit algebraic relations between the model arm's dynamic regime and the physical and control parameters of the "real" (biological, robotic, or prosthetic) arm in order to specify patterns of control parameters over time for the real arm. Saltzman and Kelso (1983a/in press) discuss two related methods for specifying these controls. Both methods are applicable to the control and coordination of artificial linkage systems (e.g., robotic or prosthetic devices), although one offers a more biologically plausible style of control than the other (see also Hogan & Cotter, 1982). The aim of both methods, however, is to make the real arm behave identically or near-identically to the model arm. Further, the essence of the task-dynamic approach lies in its account of the coordinated movement patterns that arise in a task-specific and posturally conditioned form in the model effector system. Consequently, for the purposes of the present paper, further discussion will focus on behavioral phenomena in the model articulators only. The interested reader is referred to Saltzman and Kelso (1983a/in press), however, for details concerning the hypothesized relationships between control processes of the model and real effector systems.

Task Dynamics and Speech: Bilabial gestures

The task-dynamic approach has been extended in a preliminary way to speech gestures in order to explore the hypothesis that speech production involves task-specific, dynamically specified coordination of the articulators.

Example 2: Discrete bilabial closure, unperturbed gestures. As with the limb tasks described earlier, the first step in generating simulated movements of the speech articulators is to specify the functional aspects of these gestures with reference to the movements of an effector-independent terminal device (i.e., an idealized vocal tract constriction). This is done in a two-dimensional task space whose axes represent constriction location (t_1) and constriction degree (t_2), and the topological structure of the control regime for each task-space variable is specified according to the qualitative characteristics of the given speech task. Thus, for example, discrete and repetitive speech gestures will have point attractor and limit cycle regimes, respectively, along each axis. At the task-space level, then, the control regime is an abstract one in that the constriction being controlled is independent of any particular effector system, and can refer, for example, to either a bilabial constriction produced by the lips and jaw or to a tongue-palate constriction produced by the tongue and jaw. Since simulations to date have focused on bilabial gestures, we will begin by examining a discrete bilabial closure task involving (uncoupled) point attractor dynamics along each task axis (see Figure 3A). The task-space equation of motion is expressed as follows:

$$M_T \ddot{\mathbf{t}} + B_T \dot{\mathbf{t}} + K_T \mathbf{t} = \mathbf{0}\text{, where} \qquad (6)$$

$$M_T = \begin{bmatrix} m_{T1} & 0 \\ 0 & m_{T2} \end{bmatrix}; \quad B_T = \begin{bmatrix} b_{T1} & 0 \\ 0 & b_{T2} \end{bmatrix};$$

$$K_T = \begin{bmatrix} k_{T1} & 0 \\ 0 & k_{T2} \end{bmatrix};$$

m_{T1}, m_{T2} = inertial coefficients;

b_{T1}, b_{T2} = damping coefficients;

k_{T1}, k_{T2} = stiffness coefficients.

The forms of these task-space dynamics and corresponding equation of motion are identical to those for the discrete limb reaching task described earlier (Figure 2A; Equation 1). This identity highlights the fact that functional equivalence among tasks does not depend on the specific effector systems involved, but only on the topological equivalence of dynamical regimes in the task spaces. The two main differences between the limb and speech examples are: a) the task-space axes for the bilabial task do not share a common task mass, but rather are characterized by their own inertial coefficients m_{T1} and m_{T2} (compare Equations 2 and 6); and b) the axes for the bilabial task are not differentiated into distinct "Reach" and "Normal" axes as they were in the limb reaching task. Finally, as in the reaching example, movements along the task axes do not influence one another, since the corresponding equations of motion are defined to be uncoupled.

The next step in modeling the bilabial closure is to transform the task-space system kinematically into a two-dimensional body-space system (x_1, x_2) defined in the midsagittal plane of the vocal tract and centered on the jaw's rotation axis (see Figure 3B). In contrast to the task-space regime, the body-space dynamics are effector-specific, in that they refer to the movement of a "virtual" terminal device (i.e., the bilabial constriction) of the effector system defined by the lips and jaw. The result of transforming from task-space (t_1, t_2) to jaw-space (x_1, x_2) coordinates, then, is to define a two dimensional set of motion equations, one for each axis of jaw space. As with the task-space equation, the jaw-space equation has the same form as its corresponding shoulder-space reaching equation (Equation 2). The jaw-space equation is as follows:

$$M_B \ddot{\underset{\sim}{x}} + B_B \dot{\underset{\sim}{x}} + K_B \Delta \underset{\sim}{x} = 0, \quad \text{where} \tag{7}$$

$\underset{\sim}{x}, \dot{\underset{\sim}{x}}, \ddot{\underset{\sim}{x}} = (x_1, x_2)^T$ and its derivatives with respect to time;

$\Delta\underset{\sim}{x} = \underset{\sim}{x} - \underset{\sim}{x}_0$, where $\underset{\sim}{x}_0$ = the target vector for lip protrusion (x_{01}) and lip aperture (x_{02});

$M_B = M_T$, $B_B = B_T$, and $K_B = K_T$, since no rotation is involved in the transformation from task- to jaw-space coordinates.

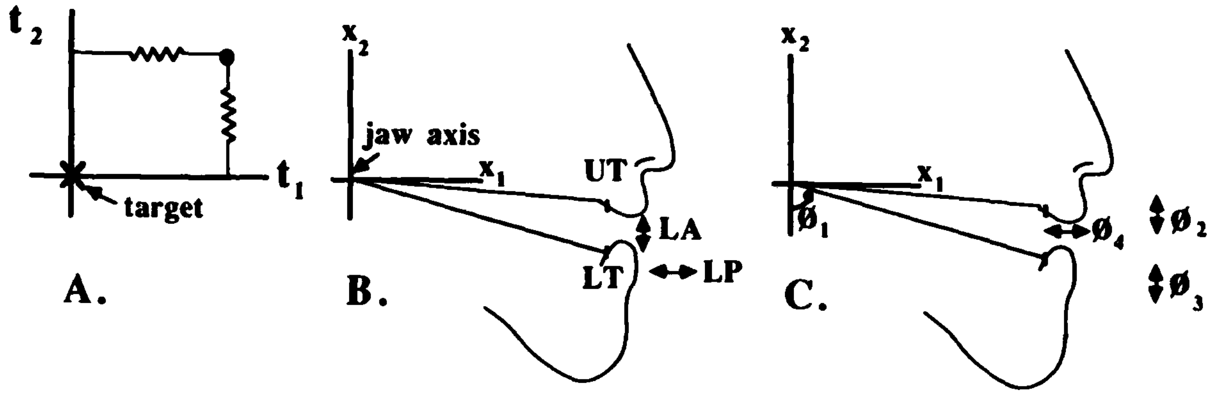


Figure 3. Bilabial tasks: A. Task space ($\underset{\sim}{t}$). Closed circle denotes current system configuration. Squiggles denote each axis' dynamics in lumped forms; B. Jaw space ($\underset{\sim}{x}$). Local tract variables (LP = lip protrusion, LA = lip aperture) are expressed in jaw coordinates. UT and LT denote positions of upper and lower front teeth, respectively; C. Model articulator space ($\underset{\sim}{\phi}$). ϕ's denote articulator variables.

Equation (7) contains a constant set of dynamic parameters, and governs the movements for the bilabial constriction along the dimensions of lip aperture (LA) and lip protrusion (LP). Lip aperture and protrusion are labeled <u>local tract variables</u>, and represent the effector-specific body-space versions of the effector-independent task-space variables of constriction degree and location, respectively. Lip aperture is defined by the vertical distance between the upper and lower lips, and lip protrusion by the horizontal distances in the anterior-posterior direction of the upper and lower lips from the upper and lower teeth, respectively. It should be noted that upper and lower lip protrusion movements are not independent in this formulation, but have been constrained to be equal in the model for purposes of simplicity. Consequently, like constriction location in task space, lip protrusion in body

space constitutes only a single degree of freedom. Finally, it should be noted that the control regimes for each jaw-space coordinate are independent, since their corresponding equations of motion are uncoupled. This is due to the fact that lip aperture and protrusion are defined parallel to the x_2 and x_1 jaw-space axes, respectively. Note, however, that such noninteracting dynamics are not usually found at the body-space level of description. For example, with movements of the tongue body orthogonal and tangential to the (curved) palate, the set of uncoupled task-space equations would be transformed into a set of (generally) coupled jaw-space equations.

The last step in modeling the closure is to transform kinematically the two dimensional jaw space regime into the coordinates of a four-dimensional model articulator space. The model articulators are moving segments that have lengths but are massless (see Figure 3C), and are defined with reference to the simplified articulatory degrees of freedom adopted in the Haskins Laboratories software articulatory speech synthesizer (Rubin, Baer, & Mermelstein, 1981). For bilabial gestures, the articulator set associated with lip aperture includes rotation of the jaw (ϕ_1), and vertical displacements of the upper lip (ϕ_2) and lower lip (ϕ_3) relative to the upper and lower front teeth, respectively; for lip protrusion, the articulator set includes yoked horizontal displacements in the anterior-posterior direction of the upper and lower lips (ϕ_4) relative to the upper and lower front teeth, respectively. Expressed in units of linear acceleration, the model articulator equation has the same form as Equation 5 and is expressed as follows (note: the angular acceleration terms in the jaw's motion equation have been multiplied by a unit scaling factor to ensure dimensional homogeneity along all articulatory degrees of freedom):

$$\ddot{\phi} + J^* M_B^{-1} B_B J \dot{\phi} + J^* M_B^{-1} K_B \Delta x(\phi) + J^* V \dot{\phi}_p = 0, \quad \text{where} \qquad (8)$$

M_B, B_B, K_B are the same constant matrices used in Equation (7); and

$\Delta x(\phi) = x(\phi) - x_0$, where $x(\phi)$ is expressed as a function of model articulator variables, and x_0 is the same constant vector used in Equation (7);

J, V, and J*: the elements of the Jacobian matrix (J, and hence also V and J*) reflect the geometrical relationships among motions of the (simplified) model speech articulators (4 degrees of freedom) and motions of the corresponding local tract variables (2 degrees of freedom); and

A: the articulator weighting matrix (A) is a component of the pseudoinverse J*. A's elements reflect task-specific constraints on the relative motions of the articulators during the closing gesture.

Given a fixed set of tuning parameters (i.e., M_T, B_T, K_T, x_0, and A) and a set of initial conditions (ϕ_I, $\dot{\phi}_I$, and hence a corresponding x_I and $\dot{x}_I$) Equation 8 will generate a pattern of coordinated motion in the model speech articulators that will achieve the task goals specified for the local tract variables. For an initial configuration (ϕ_I) corresponding to open and relatively unprotruded lips, and with an initial velocity vector of zero, the coordinated articulator movements will reflect the evolving

task-specific motions of the local tract variables en route to their specified targets (x_0), with motion characteristics (e.g., speed, degree of overshoot, etc.) specified by the pattern of M_T, B_T, and K_T parameters. Assuming the system is not perturbed during its motion trajectory, the relative extents of movement for the articulators associated with lip aperture (i.e., ϕ_1, ϕ_2, ϕ_3 in Figure 3C) will be specified by the relative values of articulator weights in the associated articulatory weighting matrix, A. Figure 4 (configurations A and B) illustrates an unperturbed movement from an initially open and relatively unprotruded configuration (Figure 4A) to a closed and relatively protruded final configuration (Figure 4B). Since the articulators associated with lip aperture were weighted equally in the corresponding A-matrix, the extents of motion for these articulators were equal over the course of the gesture.

Example 3: Immediate compensation, bilabial closure, perturbed gestures. Previous dynamical accounts of coordinated actions performed by the limbs and speech articulators have posited that invariant sets of dynamic parameters could be defined at the level of articulatory degrees of freedom (e.g., Cooke, 1980; Fel'dman, 1966; Fowler, 1977; Kelso, 1977; Polit & Bizzi, 1978). Thus, for example, discrete targeting tasks of the elbow joint were modeled as damped mass-spring systems (having point attractor dynamics) where the target angle was specified by the value of the rest angle dynamic parameter. As discussed earlier, this approach implies that the task of reaching a bilabial closure target for speech is specified according to a corresponding rest-configuration parameter for the articulators. However, recent work (Abbs & Gracco, 1983; Folkins & Abbs, 1975; Kelso et al., 1984) has shown that this formulation must be modified. In particular, the Kelso et al. (1984) study demonstrated that if the jaw is retarded en route to a bilabial closure target for /b/, then the closure is still attained and the final articulatory configuration for the perturbed movement is different from the final configuration for unperturbed movements. Significantly, the upper lip compensation is absent if the jaw is perturbed en route to an alveolar closure target for /z/. These results show that an invariant dynamic description of a movement does not apply at the articulator level, since the articulatory-dynamic parameters must be able to change according to a movement's context in an utterance-specific (i.e., /b/ vs. /z/) manner. Furthermore, the speed of these compensatory behaviors suggests that they must occur "automatically" without reference to traditional stimulus-response reaction-time correction procedures.

The task-dynamic model handles such immediate compensation as follows. Bilabial closing gestures are simulated as discrete movements toward target constrictions, using point attractor dynamics for the local tract variables of lip aperture and protrusion (see Equation 7 above). When the simulated jaw is "frozen" in place during the closing gesture at the level of the model effector system, the main qualitative features of the perturbation data are captured, in that: a) compensation is immediate in the upper and lower lips to the jaw perturbation, i.e., the system does not require reparameterization in order to compensate; and b) the target bilabial closure is reached (although with different final articulator configurations and, hence, different jaw-space locations for the closure) for both perturbed (Figure 4C) and unperturbed (Figure 4B) "trials."

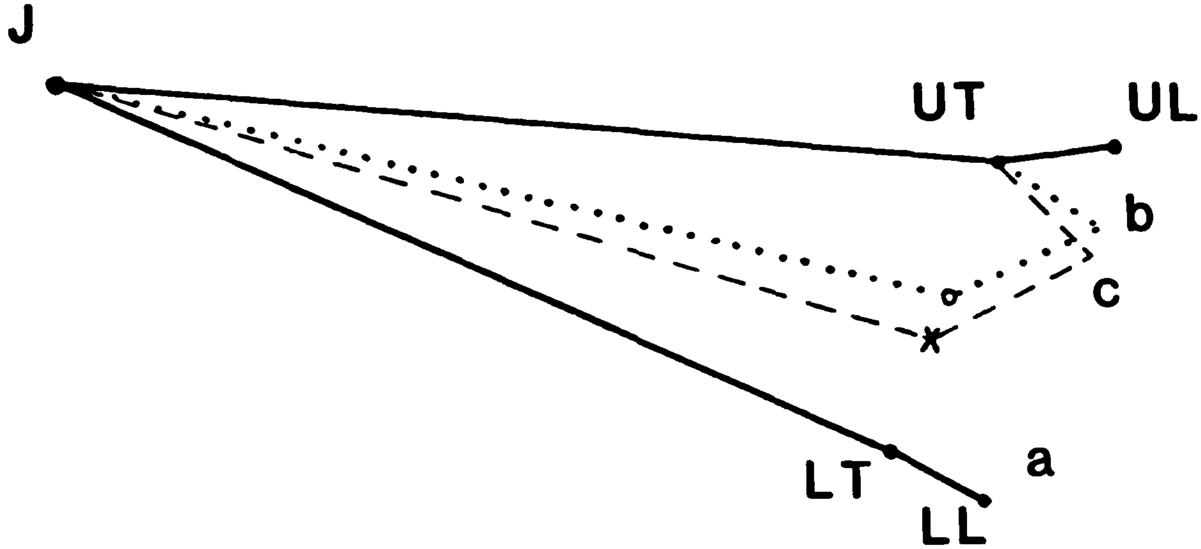


Figure 4. Simulated articulator configurations for bilabial closure task. A. Initial configuration (solid lines); B. Final configuration, unperturbed trajectory (dotted lines); C. Final configuration, perturbed trajectory (broken lines). Note that closure occurs lower in jaw space in C than in B. J = jaw axis, UT = upper teeth, UL = upper lip, LT = lower teeth, LL = lower lip.

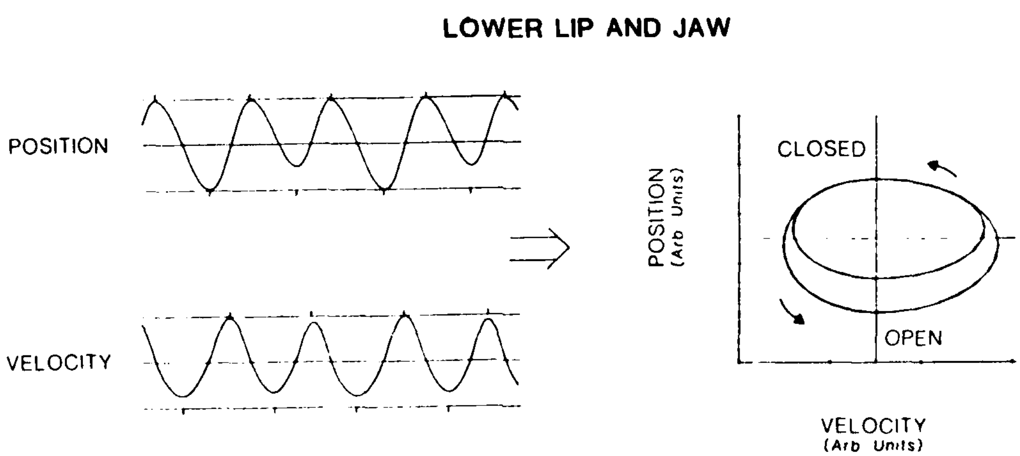


Figure 5. Simulated trajectories for lower lip height (i.e., jaw and lower lip) in the time domain (left) and phase plane (right) for a repetitive sequence of /ma/'s with alternating stress (from Kelso et al., 1985).

Example 4: Cyclic bilabial motion, unperturbed gestures. The point attractor task-space (and local tract variable) topology that was used in the discrete bilabial closure task is inappropriate, however, for generating cyclic bilabial gestures, e.g., a sequence of repeated /ma/'s as in "...mamama...". The task-dynamic model has been used to simulate a repetitive gestural sequence that is characterized by an alternating stress pattern, e.g., "...mamamama...," where the underlining denotes the pattern of stress (Browman et al., 1984). Mass-spring dynamics[1] were specified for the local tract variables of lip aperture and protrusion in order to generate sustained cyclic motions of the model articulators. Focusing on lip aperture, the parameters of rest position and stiffness were estimated from articulatory movement data collected in an experiment on reiterantly produced speech (Kelso, V.-Bateson, Saltzman, & Kay, 1985). In reiterant speech, talkers substitute a given syllable (e.g., /ma/) for the real syllables in an utterance while maintaining the utterance's normal stress pattern (e.g., the sentence "When the sunlight strikes raindrops in the air" becomes "ma ma ma ma ma ma ma ma ma ma"). The lip aperture parameters for the task-dynamic simulation were estimated using the average amplitudes and frequencies of the articulatory data obtained for the stressed and unstressed syllables spoken reiterantly at a given rate. Figure 5 illustrates the resultant cyclic trajectories for lower lip height, both in the time domain and the phase plane. For a given simulated cyclic gesture (closure-to-closure), the equilibrium position was set only once because, in the data, the jaw-lip complex returned roughly to the same position at closure for each syllable. The values for the equilibrium positions in temporally adjacent cycles alternated in value, however, since stressed syllables were found to involve greater movement amplitudes than unstressed syllables. Additionally, because closing gestures were faster than opening gestures in these data, two values of stiffness were specified within each cycle: one at the start of the opening gesture and another at the start of the closing gesture. The set of task-dynamic parameters were invariant, therefore, over the course of a given opening or closing gesture.

Conclusion

The task-dynamic model is able to generate coordinated movement patterns for the model articulators in both discrete and cyclic unperturbed (bilabial) utterances. Additionally, for discrete bilabial closing gestures it provides task-specific patterns of compensatory responses to jaw perturbations that are qualitatively similar to those observed experimentally. Finally, Browman et al. (1984) have used sets of simulated articulator trajectories from an alternating stress, repetitive, bilabial speech task as inputs to the Haskins Laboratories articulatory speech synthesizer (Rubin et al., 1981; see also Example 2 above) with promising acoustic and perceptual results. Note that, although these simulated utterances involve a simple stress pattern and segmental structure, the task-dynamic approach to articulatory speech synthesis could certainly be used to generate more complex utterances on a gesture-by-gesture basis. The elegance of the procedure would still be maintained, however, since utterance-specific and contextually variable patterns of articulator trajectories and compensatory responses would still emerge automatically as implicit consequences of task space control regimes that are invariant within a given speech gesture. There is no need to invoke either explicit trajectory planning or replanning procedures on a timeframe-to-timeframe basis within the gesture. My colleagues and I are encouraged by these preliminary results, and are currently engaged in

extending the task-dynamic model to account for phenomena such as coarticulation (e.g., Harris, 1984; Kent & Minifie, 1977) and relative timing (e.g., Kent, Carney, & Severeid, 1974; Tuller, Kelso, & Harris, 1982, 1983) among serially ordered speech gestures.

References

Abbs, J. H., & Gracco, V. L. (1983). Sensorimotor actions in the control of multimovement speech gestures. Trends in Neurosciences, 6, 393-395.

Abraham, R. H., & Shaw, C. D. (1982). Dynamics--The geometry of behavior. Santa Cruz, CA: Aerial.

Arbib, M. A. (1984). From synergies and embryos to motor schemas. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. Amsterdam: North-Holland.

Benati, M., Gaglio, S., Morasso, P., Tagliasco, V., & Zaccaria, R. (1980). Anthropomorphic robotics. II. Analysis of manipulator dynamics and the output motor impedance. Biological Cybernetics, 38, 141-150

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon.

Bizzi, E., & Abend, W. (1982). Posture control and trajectory formation in single and multiple joint arm movements. In J. E. Desmedt (Ed.), Brain and spinal mechanisms of movement control in man: New developments and clinical applications. New York: Raven.

Bizzi, E., Accornero, N., Chapple, W., & Hogan, N. (1981). Processes underlying arm trajectory formation. In C. Ajmone-Marsan & O. Pompeiano (Eds.), Brain mechanisms of perceptual awareness and purposeful behavior (pp. 311-318). IBRO Monograph Series. New York: Raven.

Boylls, C. C., Jr., & Greene, P. H. (1984). Introduction: Bernstein's significance today. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. Amsterdam: North-Holland.

Browman, C. P., & Goldstein, L. M. (1985). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), Phonetic linguistics. New York: Academic.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984). Articulatory synthesis from underlying dynamics. Journal of the Acoustical Society of America, 75, S22-S23. (Abstract)

Cooke, J. D. (1980) The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Delatizky, J. (1982). Final position control in simulated planar horizontal arm movements. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Easton, T. A. (1972). On the normal use of reflexes. American Scientist, 60, 591-599.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic analysis of execution by man of the simplest motor tasks. Biophysics, 11, 766-775.

Fitch, H., & Turvey, M.T. (1978). On the control of activity: Some remarks from an ecological point of view. In D. Landers, & R. Christina (Eds.), Psychology of motor behavior and sports. Urbana, IL: Human Kinetics.

Folkins, J. W., & Abbs, J.H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. Journal of Speech and Hearing Research, 18, 207-220.

Fowler, C. (1977). Timing control in speech production. Indiana University Linguistics Club, Bloomington, IN.

Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic.

Greene, P. H. (1972). Problems of organization of motor systems. Progress in Theoretical Biology, 2, 303-338.

Harris, K. S. (1984). Coarticulation as a component of articulatory descriptions. In R. G. Daniloff (Ed.), Articulation assessment and treatment issues. San Diego, CA: College Hill Press.

Hebb, D. O. (1949). The organization of behavior. New York: Wiley.

Hogan, N., & Cotter, S.L. (1982). Cartesian impedance control of a nonlinear manipulator. In W. J. Book (Ed.), Robotics research and advanced applications (pp. 121-128). New York: ASME.

Hollerbach, J.M. (1982). Computers, brains, and the control of movement. Trends in Neurosciences, 5, 189-192

Hollerbach, J.M., & Atkeson, C. G. (in press). Characterization of joint-interpolated arm movements. Experimental Brain Research Supplementum.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 49-70). New York: North-Holland.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., Tuller, B., V-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation adaptation to jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Perception and Performance, 10, 812-832.

Kelso, J. A. S, V-Bateson, E., Saltzman, E. L., & Kay, B. A. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Kent, R. D., Carney, P. J., & Severeid, L. R. (1974). Velar movement and timing: Evaluation of a model for binary control. Journal of Speech and Hearing Research, 17, 470-488

Kent, R. D., & Minifie, F. D. (1977). Coarticulation in recent speech production models. Journal of Phonetics, 5, 115-133.

Klein, C. A., & Huang, C-H. (1983.) Review of pseudoinverse control for use with kinematically redundant manipulators. IEEE Transactions on Systems, Man, and Cybernetics, SMC-13,245-250.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A.S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination (pp. 5-78). Chichester: John Wiley.

Lashley, K. S. (1930). Basic neural mechanisms in behavior. Psychological Review, 37, 1-24

Marsden, C. D., Merton, P. A., & Morton, H. B. (1983). Rapid postural reactions to mechanical displacement of the hand in man. In J. E. Desmedt (Ed.), Motor control mechanisms in health and disease. New York: Raven.

Morasso, P. (1981). Spatial control of arm movements. Experimental Brain Research, 42, 223-227.

Nashner, L. M., & McCollum, G. (1985). The organization of human postural movements: A formal basis and experimental synthesis. The Behavioral and Brain Sciences, 8, 135-150.

Polit, A., & Bizzi, E. (1978). Processes controlling arm movements in monkeys. Science, 201,1235-1237.

Rubin, P., Baer, T., & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. Journal of the Acoustical Society of America, 70, 321-328.

Saltzman, E. (1979). Levels of sensorimotor representation. Journal of Mathematical Psychology, 20, 91-163.

Saltzman, E., & Kelso, J. A. S. (1983a/in press). Skilled actions: A task dynamic approach. Haskins Laboratories Status Report on Speech Research, SR-76, 3-50. Also Psychological Review, in press.

Saltzman, E., & Kelso, J. A. S. (1983b). Toward a dynamical account of motor memory and control. In R. Magill (Ed.), Memory and control of action(pp. 17-38). Amsterdam: North-Holland.

Saltzman, E., & Kelso, J. A. S. (1985). Synergies: stabilities, instabilities, and modes. The Behavioral and Brain Sciences, 8, 161-163.

Soechting, J. E., & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. Journal of Neuroscience, 1, 710-720.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1983). Further evidence for the role of relative timing in speech: A reply to Barry. Journal of Experimental Psychology: Human Perception and Performance, 9, 829-833.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

Wadman, W. J., Denier van der Gon, J. J., & Derksen, R. J. A. (1980). Muscle activation patterns for fast goal-directed arm movements. Journal of Human Movement Studies, 6, 19-37.

Whitney, D. E. (1972). The mathematics of coordinated control of prosthetic arms and manipulators. ASME Journal of Dynamic Systems, Measurement and Control, 94, 303-309.

Footnote

[1]Mass-spring and limit cycle dynamics can produce near-identical motion trajectories in the absence of perturbations to the system. Since, the modeled cyclic bilabial gestures were unperturbed by design, (undamped) mass-spring dynamics were adequate for these purposes. The model is presently being extended, however, to include limit cycle dynamics at the task-space level, in order to explore the simulated effects of perturbations introduced during cyclic speech tasks.

SOME OBSERVATIONS ON THE DEVELOPMENT OF ANTICIPATORY COARTICULATION*

Bruno H. Repp

Abstract. The influence of vowel quality on various temporal and spectral properties of preceding acoustic segments was investigated in utterances containing [ə#CV] sequences produced by two girls aged 4;8 and 9;5 years and by their father. The younger (but not the older) child's speech showed a systematic lowering of [s] noise and [t^h] release burst spectra before [u] as compared to [i] and [æ]. The older child's speech, on the other hand, showed an orderly relationship of the second-formant frequency in [ə] to the transconsonantal vowel. Both children tended to produce longer [s] noises and voice onset times as well as higher second-formant peaks at constriction noise offset before [i] than before [u] and [æ]. All effects except the first were shown by the adult who, in addition, produced first-formant frequencies in [ə] that anticipated the transconsonantal vowel. These observations suggest that different forms of anticipatory coarticulation may have different causes and may follow different developmental patterns. A strategy for future research is suggested.

The development of coarticulation in children's speech production is a topic of great current interest, although data are still scarce. It is commonly assumed that children coarticulate less than adults, especially with regard to anticipatory effects that are said to be planned, and there is some preliminary evidence from acoustic analyses and from physiological studies to support this notion (see Kent, 1983). A reduction in the extent of coarticulation is taken to reflect an underlying general tendency toward producing speech segment by segment, which decreases with age (Kent, 1983).

In the present pilot study, acoustic measures of several anticipatory coarticulation effects were obtained from two children and their father. Because of this small sample size, the data are intended to stimulate further research rather than to establish firm developmental patterns. Nevertheless, the familial relatedness of the three subjects may have reduced irrelevant individual differences, thus lending the data somewhat more generality than a sample of three unrelated individuals would have provided.

I. Methods

A. Subjects

The subjects were two sisters aged 4;8 and 9;5 years and their father (the author). The children are monolingual speakers of American English; the

---

*Journal of the Acoustical Society of America, in press.
Acknowledgment. This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. I am grateful to Catherine Best, Sarah Hawkins, Joanne Miller, Susan Nittrouer, Daniel Recasens, Sigfrid Soli, Michael Studdert-Kennedy, Douglas Whalen, and Grace Yeni-Komshian for helpful comments on earlier drafts.

adult is a native speaker of German who speaks English almost exclusively, though not without an accent.

B. Utterances and Procedure

Each subject produced six words, sea, sand, soup, tea, tan, and tooth, five times in the carrier phrase "I like the ...". The children repeated each sentence after their father, taking turns at speaking first.[1] The recordings were made in a sound-attenuated booth, with all three talkers facing a single microphone.

C. Acoustic Analysis

The children's utterances were low-pass filtered at 9.6 kHz and digitized at a 20 kHz sampling rate with high-frequency pre-emphasis. A 24-coefficient LPC analysis with automatic peak-picking and subsequent hand-editing of inconsistencies yielded estimates of formant frequencies. A numerical index of the relative high-frequency content of the spectrum in a given 20-ms analysis frame was provided by the first LPC reflection coefficient, which is the (negative, normalized) average of the cosine-weighted spectrum (see Markel & Gray, 1976). Temporal measures were obtained from oscillographic displays. Means and standard deviations of the various measures were computed across the five tokens of each utterance. The adult's utterances were analyzed similarly, using a 10 kHz sampling rate for digitization and a 14-coefficient LPC model.

II. Results

A. Effects of Vocalic Context on Voiceless Interval Durations

Table 1 shows two coarticulatory effects in the temporal domain: [s] noise durations were longest before [i] and shortest before [æ], and [t^h]

Table 1

Means and Standard Deviations (in Parentheses) of Some Voiceless Segment Durations (ms).

	Child A (4;8 yrs)	Child B (9;5 yrs)	Adult
[s(V)] fricative noise			
V = [i]	232 (24)	222 (34)	228 (9)
V = [æ]	184 (25)	189 (21)	173 (9)
V = [u]	207 (27)	202 (17)	197 (9)
[t^h(V)] burst + aspiration (VOT)			
V = [i]	90 (16)	107 (5)	76 (10)
V = [æ]	75 (12)	89 (10)	64 (15)
V = [u]	84 (21)	84 (16)	50 (7)

burst plus aspiration (i.e., acoustic voice onset time or VOT) was longest before [i] also. In separate one-way analyses of variance, the [s] duration differences reached significance for the younger child, $F(2,12) = 4.47$, $p = .0354$, while the VOT differences reached significance for the older child, $F(2,12) = 6.24$, $p = .0139$. Both effects were highly significant in the adult, $F(2,12) = 59.0$, $p < .0001$, and $F(2,12) = 7.35$, $p = .0083$, respectively. All three talkers showed similar patterns, however, and the lower reliability of the children's results may be attributed to their greater variability (cf. Smith, Sugarman, & Long, 1983).[2]

B. Effects of Vocalic Context on Constriction Noise Spectra

A lowering of [s] frication and [t^h] release burst spectra due to anticipatory lip rounding for [u] has been observed in adults (Mann & Repp, 1980; Sereno, Baum, Marean, & Lieberman, 1985; Soli, 1981; Turnbaugh, Hoffman, Daniloff, & Absher, 1985; Zue, 1976). Visual inspection of average [s] noise offset and [t^h] burst onset spectra (both representing noise immediately preceding the release of the constriction) revealed a clear shift of the energy maximum towards lower frequencies (5-6 kHz) before [u] as compared to [i] and [æ] (around 8 kHz) in the younger child. Neither the older child nor the adult showed such a shift.

To gain statistical support for these observations, and to examine the time course of the effect in the [s] noise, analyses of variance were conducted on the average first LPC reflection coefficients obtained for three (slightly overlapping) consecutive 60-ms segments of the [s] noises of each talker. For the younger child, there were highly significant effects of vocalic context, $F(2,12) = 14.22$, $p = .0007$, and of time, $F(2,24) = 19.80$, $p < .0001$, as well as a two-way interaction, $F(4,24) = 5.56$, $p = .0026$. The coarticulatory effect increased with proximity to the vowel but was clearly present throughout the fricative noise. The older child, on the other hand, showed no significant effects, even though spectral variability was lower. The adult talker also showed significant effects of vocalic context, $F(2,12) = 9.89$, $p = .0029$, and of time, $F(2,24) = 5.98$, $p = .0078$, but the pattern was different: the average [s] spectra were lowest before [æ] and highest before [u]; moreover, these differences resided mainly between 1-3 kHz.

The noise spectra were also examined for peaks in the second-formant (F2) region that anticipate F2 in the following vowel, a lingual coarticulation effect that is distinct from the global spectral shifts due to anticipatory lip rounding (see Soli, 1981). F2 frequency estimates derived from the 20-ms LPC analysis frames closest to [s] noise offset and [t^h] burst onset are reported in Table 2. There was a significant effect of vocalic context for the younger child, $F(2,24) = 11.28$, $p = .0004$: In both [s] offset and [t^h] onset spectra, F2 was highest preceding [i]. The older child, despite more pronounced F2 peaks and lower variability, showed only a nonsignificant tendency in the same direction, $F(2,24) = 3.32$, $p = .0531$. The adult's F2 peaks were significantly higher before [i] than before [u], $F(1,16) = 50.36$, $p < .0001$; before [æ], no reliable F2 peaks could be found (cf. Soli, 1981).

C. Effects of Vocalic Context on [ə] Formant Frequencies

Vowel-to-vowel anticipatory coarticulation across an intervening consonant has been observed in adults, especially in [ə] (Alfonso & Baer, 1982; Fowler, 1981). Table 2B shows means and standard deviations of F2

Table 2

Means and Standard Deviations (in Parentheses) of F2 Frequencies at [s] Noise Offset, at [t^h] Burst Onset, and in the Preceding [ə] (Hz).

(A)	Child A (4;8 yrs)	Child B (9;5 yrs)	Adult
at [s(V)] *noise offset*			
V = [i]	3241 (168)	2385 (92)	1957 (66)
V = [æ]	2899 (186)	2331 (120)	----------
V = [u]	2866 (159)	2203 (90)	1547 (51)
at [t^h(V)] burst onset			
V = [i]	3176 (127)	2492 (144)	2191 (259)
V = [æ]	2998 (63)	2357 (33)	----------
V = [u]	3050 (90)	2430 (147)	1757 (116)

(B)			
in [ə] preceding [#sV]			
V = [i]	2846 (123)	2107 (50)	1482 (26)
V = [æ]	2885 (114)	2049 (59)	1421 (15)
V = [u]	2863 (104)	2018 (64)	1490 (75)
in [ə] preceding [#t^hV]			
V = [i]	2866 (169)	2168 (55)	1467 (18)
V = [æ]	2857 (108)	2154 (24)	1418 (45)
V = [u]	2934 (52)	2077 (47)	1418 (45)

frequencies averaged over the whole voiced signal portion corresponding to [ə] in the word the as a function of following consonant and vowel. There were no systematic contextual effects for the younger child. The older child, in contrast, showed a systematic decrease of F2 as the vowel in the following syllable changed from [i] to [æ] to [u], $F(2,24) = 7.75$, $p = .0025$, as well as higher F2 frequencies preceding [t^h] than [s], $F(1,24) = 15.85$, $p = .0006$. Both effects were present throughout the [ə] vowel. The first formant (F1) did not show any significant differences for either child. The adult also showed a significant effect of vowel context on F2, $F(2,24) = 5.32$, $p = .0123$, due to elevated F2 frequencies preceding [i]. In addition, he showed an effect on F1, which was significantly higher (by about 33 Hz) preceding [æ] than preceding [i] and [u], $F(2,24) = 8.31$, $p = .0018$, thus anticipating the F1 differences between these vowels.

III. Discussion

It is not possible to derive any conclusions about general developmental trends from these limited data. Nevertheless, they may serve as a basis for formulating hypotheses about the development of anticipatory coarticulation, to be tested in the future with larger subject groups or in longitudinal studies.

Two coarticulatory effects in the temporal domain were shown by both children and by the adult, though with different degrees of reliability. One of these, the effect of the following vowel on [s] noise duration, may be due to an earlier release of the constriction preceding more open vowels (Schwartz, 1969). DiSimoni (1974) and Weismer and Elbert (1982) have obtained similar differences in preschool children. The other effect apparently shown by all three subjects was that of vowel context on VOT. Related findings in the literature (Fourakis, 1986; Klatt, 1975; Port & Rotunno, 1979; Weismer, 1979) are at least partially consistent with the longer VOTs preceding [i] observed here. These effects may have kinematic or aerodynamic causes that make them difficult to avoid at any age.

A third effect that was probably present in all three talkers, although it was not quite significant in the older child, concerns differences in the location of F2 peaks at the release of a fricative constriction or of a stop occlusion. These differences probably reflect differences in tongue body position in anticipation of the upcoming vowel (Soli, 1981), although anticipatory lip rounding may also play a role. Similar effects were found in a 3;6 year old child by Sereno et al. (1985), and in several 3- and 5-year-old children by Turnbaugh et al. (1985). This may be another obligatory effect; without any anticipation, the vowel might sound abnormally diphthongized.

By contrast, certain other coarticulatory effects may be optional and subject to developmental trends. Changes in F2 of [ə] in anticipation of the later-occurring vowel clearly were shown only by the older child and the adult. This effect probably reflects differences in tongue body position (Alfonso & Baer, 1982); note that it was not prevented by an intervening alveolar consonant that also involves the tongue (see Recasens, 1984). This relatively long-range anticipatory lingual coarticulation across an obstacle may be a skill that is acquired relatively late as a child gets acquainted with the fine details of spoken language. The same might be said about the vocalic context effect on F1 frequency in [ə], which was shown by the adult alone and may reflect anticipatory adjustments in jaw elevation. Note that,

to the extent that these articulatory postures are not maintained during the intervening consonant constriction, they must indeed be considered planned.

The most unusual finding concerns the overall weighting of constriction noise spectra. A lowered [s] noise or [t^h] release burst spectrum before rounded vowels such as [u] most likely reflects an effect of anticipatory lip rounding, although changes in tongue body position could also play a role (Carney & Moll, 1971). Such an effect was observed very clearly in the younger child but not in the older child, and it was reversed in the adult. While the reversal may be atypical (it could reflect back cavity resonances brought into play by leaky [s] constrictions characteristic of this adult speaker), it is interesting to note that Nittrouer (1985), in a recent thorough developmental study, has observed that fricative-vowel coarticulation (in terms of global spectral shifts in the noise) does decline with age. The present data are consistent with such a trend, even though its reasons are far from clear at present.

IV. Conclusions

The various patterns of results observed in this pilot study suggest that phenomena commonly lumped together under the heading of coarticulation may have diverse origins and hence different roles in speech development. Some forms of coarticulation are an indication of advanced speech production skills whereas others may be a sign of articulatory immaturity, and yet others are neither because they simply cannot be avoided. Therefore, it is probably not wise to draw conclusions about a general process called coarticulation from the study of a single effect. Indeed, such a general process may not exist. It is suggested that future research adopt the multi-pronged approach illustrated by this pilot study to examine the interrelationships among diverse coarticulatory phenomena, their individual causes, and their patterns of development.

References

Alfonso, P. J., & Baer, T. (1982). Dynamics of vowel articulation. *Language and Speech*, *25*, 151-173.

Carney, P. J., & Moll, K. L. (1971). A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica*, *23*, 193-202.

DiSimoni, F. G. (1974). Influence of vowel environment on the duration of consonants in the speech of three-, six-, and nine-year-old children. *Journal of the Acoustical Society of America*, *55*, 360-361.

Fourakis, M. (1986). A timing model for word-initial CV-syllables in Modern Greek. *Journal of the Acoustical Society of America*, *79* (in press).

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research*, *40*, 127-139.

Kent, R. D. (1983). The segmental organization of speech. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 57-90). New York: Springer-Verlag.

Klatt, D. H. (1975). Voice-onset-time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, *18*, 686-706.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, *28*, 213-228.

Markel, J. D., & Gray, A. H., Jr. (1976). *Linear prediction of speech*. Berlin: Springer-Verlag.

Nittrouer, S. (1985). The role of coarticulation in the perception and production of speech by young children (3 to 7 years). Unpublished doctoral dissertation, City University of New York.

Port, R. F., & Rotunno, R. (1979). Relation between voice-onset time and vowel duration. Journal of the Acoustical Society of America, 66, 654-662.

Recasens, D. (1984). Vowel-to-vowel coarticulation in Catalan VCV sequences. Journal of the Acoustical Society of America, 76, 1624-1635.

Schwartz, M. F. (1969). Influence of vowel environment upon the duration of /s/ and /š/. Journal of the Acoustical Society of America, 46, 480-481.

Sereno, J. A., Baum, S. R., Marean, G. C., & Lieberman, P. (1985). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. Journal of the Acoustical Society of America, 77 (Suppl. No. 1), S7. (A)

Smith, B. L., Sugarman, M. D., & Long, S. H. (1983). Experimental manipulation of speaking rate for studying temporal variability in children's speech. Journal of the Acoustical Society of America, 74, 744-749.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. Journal of the Acoustical Society of America, 70, 976-984.

Turnbaugh, K. R., Hoffman, P. R., Daniloff, R. G., & Absher, R. (1985). Stop-vowel coarticulation in 3-year-old, 5-year-old, and adult speakers. Journal of the Acoustical Society of America, 77, 1256-1257.

Weismer, G. (1979). Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. Journal of Phonetics, 7, 197-204.

Weismer, G. (1980). Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations. Journal of Phonetics, 8, 427-438.

Weismer, G. (1981). Temporal characteristics of the laryngeal devoicing gesture for voiceless consonants and fricative-stop clusters: Influences of vowel environment and speaker age. Journal of the Acoustical Society of America, 69, S68 (Abstract).

Weismer, G., & Elbert, M. (1982). Temporal characteristics of 'functionally' misarticulated /s/ in 4-to-6-year-old children, Journal of Speech and Hearing Research, 25, 275-287.

Zue, V. W. (1976). Acoustic characteristics of stop consonants: A controlled study. Lincoln Laboratory Technical Report No. 523 (Lexington, MA).

Footnotes

[1]Apart from overall timing and intonation, it seems unlikely that the children directly mimicked any phonetic features of the adult's productions. Rather, it is assumed that the children generated their utterances from lexical representations of the (known) target sentences.

[2]The effects of vowel context on [t^h] closure duration and on the total [t^h] voiceless interval seemed less systematic. In a combined analysis of [s] and total [t^h] durations, however, none of the talkers showed a significant consonant x vowel interaction, so that the effect of vowel context on the two voiceless interval durations may have been similar (cf. Weismer, 1981). It might also be noted that the average durations of the [s] and [t^h] voiceless intervals were virtually identical in all three talkers (cf. Weismer, 1980).

THE ROLE OF PRODUCTION VARIABILITY IN NORMAL AND DEVIANT DEVELOPING SPEECH*

Katherine S. Harris,† Judith Rubin-Spitz,† and Nancy S. McGarr†

The idea of an underlying structure that is given some kind of imperfect surface manifestation is, of course, a rather common one in description of behavioral phenomena in general, and linguistic systems in particular. Following the lead of Jacobson's (1968) famous monograph investigations of child language have been couched in terms of underlying phonological systems, related to a child's phonetic output by rewrite rules, like the rules governing morphophonemic alternations in adult speech. Thus, a child who omits the final /g/ in the word "dog," but will produce the diminutive "doggie" may be described as having an underlying representation that includes the /g/, with a rule that deletes it in syllable-final position.

Many scholars, notably Smith (1973) and Ingram (1976), have asserted that the underlying phonology of normal children at the time of beginning vocabulary development is that of the ambient community. This belief rests in part on old anecdotal evidence that children often can recognize words that they cannot produce, and in part on more recent evidence regarding the ability of infants to discriminate differing speech sounds (Eimas, 1982). However, as Studdert-Kennedy points out (1985) "I do not doubt that infants can form auditory categories, but there is no evidence that this capacity is either needed for or brought to bear on early speaking."

Much the same view of the relationship of two levels is often taken of the underlying phonology in functionally misarticulating children (for a history of the use of phonological process analysis within speech pathology, see Edwards & Shriberg, 1983). That is, it has often been assumed that the misarticulating child has a normal underlying perceptual process, but obeys rule-governed restrictions in output.

Recently, Elbert, Dinssen and Weismer (1984) and Maxwell (1979) have suggested that misarticulating children differ among themselves in the

---

*In J. L. Lauter (Ed.), Proceedings of the Conference on the Planning and Production of Speech in Normal and Hearing-impaired Individuals (ASHA Reports 15). Rockville, MD: ASHA, 1985.

†Also Center for Research in Speech and Hearing Sciences, Graduate School and University Center, City University of New York.

Acknowledgment. A portion of this research was conducted in partial fulfillment of the requirements for the Ph.D. at The City University of New York by Judith Rubin-Spitz. This, and the other research cited, was supported by Grants NS-13617 and HD-01994 to Haskins Laboratories, and Grant NS-17764 to the Graduate School and University Center, City University of New York, as well as the Faculty Research Award Program, City University of New York.

relationship of underlying and surface forms. While some children give evidence, either by the presence of morphonemic alternations (e.g., /dɔ/ but /dɔgI/) or by preservation of acoustic differences in output for two forms in which a phone is omitted in transcriptional description, others do not. These authors suggest, therefore, that the nature of a child's phonological structure should be demonstrated on a phone-by-phone basis, rather than assumed.

It is possible to take the more radical position that description of children's early word attempts might be couched in auditory and motoric rather than linguistic terms (Studdert-Kennedy, 1985). After all, it is not necessary to assume that the child has internalized phonological categories that conform to the description of adult linguistic behavior (Harris, 1983; Menn, 1980; Menyuk & Menn, 1979). The fact that transcription has been the method of choice for describing children's production has tended to push description towards adult categories. However, Ferguson has presented evidence that early words are learned on a one-by-one basis (Ferguson & Farwell, 1975) and that attempts at an early word are highly variable. While it is extremely difficult to abandon the transcriptional description of words, even transcriptions show that ubiquitous variability is an essential component of the description of the child's categories.

This same variability has been repeatedly shown in instrumental descriptions as characteristic of the speech of children, even when they produce apparently mature forms (Kent, 1976). Eguchi and Hirsh (1969) described the spectral variability of production of vowels in children's speech. While the extent to which their data were affected by measurement error has been the subject of some discussion (Monsen & Engebretson, 1983), there seems to be little doubt about the appropriateness of Eguchi and Hirsh's characterization of the variability phenomenon itself. Similar production variability has been shown to characterize temporal aspects of developing speech production capabilities (see, e.g., Smith, 1978)

We emerge, then, from the description of normal child phonology with two general principles. First, a phonological inventory description must be supported by production data of some sort that demonstrates the differentiation of units that are presumed to be phonologically distinct. Often, forms distinct in the adult model are collapsed in the child's output, or are differentiated on a basis that is different from the adult. Second, it may be that the description of a child's speech in terms of an underlying phonological structure fails to capture at least the important variability aspect of performance.

When we turn to deaf children, we find that the same kind of phonological structure approach has been used in describing their speech, especially by Monsen (1976, 1983) and by Fisher, King, Parker, and Wright (1983). For hearing-impaired children there is, of course, no question that the representations supporting the phonological structure must be very different from that of the hearing community, since we presume that the sensory information on which such children base any structure and maintain differentiation between items is very different from that for normals. Thus, in Fisher et al.'s description, a single form is produced by deaf children for forms that are differentiated in the adult model, or a given contrast, while preserved, is preserved in phonetically different terms. One of the most interesting points made by Fisher and his colleagues (op. cit.) is that

intelligibility for those deaf speakers who maintain a system of deviant contrasts may be reduced by a speech training regime that moves some phones towards the normal model, but removes certain contrasts that are preserved on a deviant basis.

What kind of evidence might be marshalled in support of the point of view that the oral deaf preserve contrasts between phones as normals do? We can examine, carefully and systematically, the variability of production of some class of sounds. A deviant phonology would be indicated by normal production variability, co-occurring with a failure to differentiate pairs of sounds, or an abnormally based distinction.

An indirect form of evidence for the "deviant phonology" hypothesis could be provided by the listener effect, investigated by several researchers at the Central Institute for the Deaf. If deaf speakers differentiate between sounds in production in a way that is different from normal, then teachers experienced in listening to deaf speakers might be able to invoke a special listening strategy, based on the use of cues that naive listeners ignore. For example, if it were true that some deaf speakers systematically substitute fundamental frequency variation for formant variation (Angelocci, Kopp, & Holbrook, 1964), then an experienced listener might simply focus on this characteristic as a way of differentiating vowels (or classes of vowels). The listeners would then show a heavier dependence on F_0 than on spectral characteristics of individual tokens. Alternatively, if deaf speakers simply overlay some abnormal characteristic (Stevens, Nickerson, & Rollins, 1983), such as too high or too low pitch on their speech, experienced listeners might learn to ignore the deviant overlay, and focus on vowel cues. In this case, the pattern of differentiation would be the same for experienced and inexperienced listeners, although experienced listeners would show superior performance.

An essential component of the listener effect is that listeners must be able to identify speakers as deaf. Some time ago, Calvert (1961) demonstrated very convincingly that experienced teachers of the deaf can identify speakers as deaf, but that the teachers' performance depends very heavily on the evidence of articulator movement in the samples judged--that is, the time-dependent deviance of deaf articulatory patterns is detectable, and hence, might serve as the basis of a detection strategy. Moreover, the fact that sustained vowels produced by deaf talkers are less readily identified than vowels produced in context suggests that such identification does not depend on an overlaid characteristic, such as voice quality.

In what follows, we will discuss three studies that bear on the issues above. The first is an unpublished doctoral dissertation by Judith Rubin (1984). Obviously, there is a great deal more detail in her study than can be reported here. We will then go on to discuss some physiological work on interarticulator timing in the productions of deaf talkers (McGarr & Gelfer, 1983; McGarr & Harris, 1983; McGarr & Löfqvist, 1982) and also in normal speakers (Harris, Tuller, & Kelso, 1985; Tuller & Kelso, 1984; Tuller, Kelso, & Harris, 1982, 1983).

The object of Rubin's study was, first, to make a direct test of the hypothesis that deaf speakers produce vowels with the same variability as normal talkers. Beyond that, she wanted to compare the strategies that experienced and inexperienced listeners use in decoding deaf and normal vowels.

The subjects of her study were six orally trained, severely or profoundly hearing-impaired high school students and two age-matched normals. The speakers were asked to say "You got me the bVb" with any of seven test vowels in the vowel slot. Each token was produced 15 times. The results were analyzed acoustically, using an LPC algorithm; F_0, F_1, F_2 and duration were measured.

In the perceptual part of the study, experienced and inexperienced listeners were asked to make two judgments--first, they were asked to identify whether each vowel token was produced by a deaf or a normal talker. Second, they were asked to identify the vowel. Stimuli were presented in three conditions--first, the whole utterance; second, the /bVb/ syllable alone; and third, a short, more-or-less steady state segment gated out of the middle of the /bVb/. The stimuli were grouped by condition, but not by speaker.

We will first describe the results of the acoustic formant analysis. First, as has been previously reported (Monsen 1976) on average, deaf talkers show a reduced range of average F_1 and F_2 values, relative to normals--durations are prolonged as has been previously reported, and fundamental frequency is a little higher on average. (Note that the talkers were preselected to avoid subjects with such severe source problems that LPC analysis would become problematic.) However, when we look at individual talkers, comparing mean plots and variability plots, a more complicated picture emerges.

While individual differences are not discussed here in detail, some of the speakers showed small variability for the point vowels (/i/, /a/, and /u/), with much greater variability for intermediate vowels such as /e/. Some showed overlap between front and back vowels while some showed a great deal of variability for all vowels. Thus the placement of the average values in F_1-by-F_2 space does not predict the relative variability of the tokens around average values.

This point is illustrated in the average data for two hearing-impaired speakers. Average vowels for the first speaker shown in Figure 1 are quite appropriately distributed in formant space.

In Figure 2, the ranges of the tokens for the same speaker are shown by adding lines drawn to enclose the points representing all tokens. For this speaker, the three point vowels /i, a, u/ are reasonably well defined; however, intermediate vowels are much more variable.

Average values for a second deaf speaker are very similar to those for the first, as shown in Figure 3, but when we examine the distribution around the average values, as shown in Figure 4, we find a great deal of smear for all vowels. That is, the average values do not give a clear picture of the token-to-token variability.

Figure 5 shows the standard deviations of F_1 and F_2 for the six talkers, while Figure 6 shows standard deviations for the four acoustic measures summarized in a somewhat different fashion. The important point here is that deaf talkers are statistically significantly more variable than normals on every acoustic dimension. Thus, a description of average formant values fails to capture the characteristics of their vowel systems.

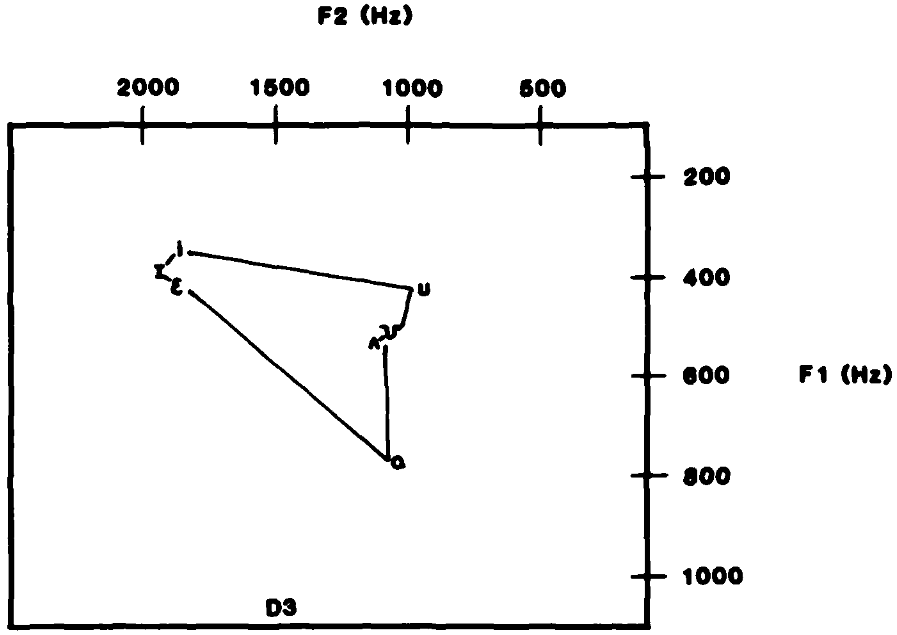


Figure 1. Average vowels for Talker D3.

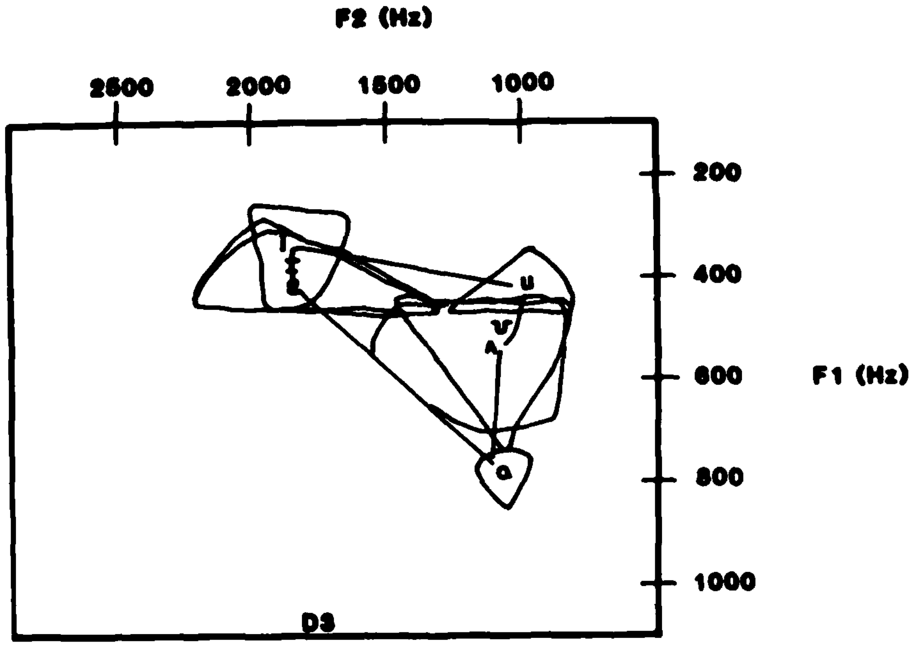


Figure 2. Range of vowels for Talker D3.

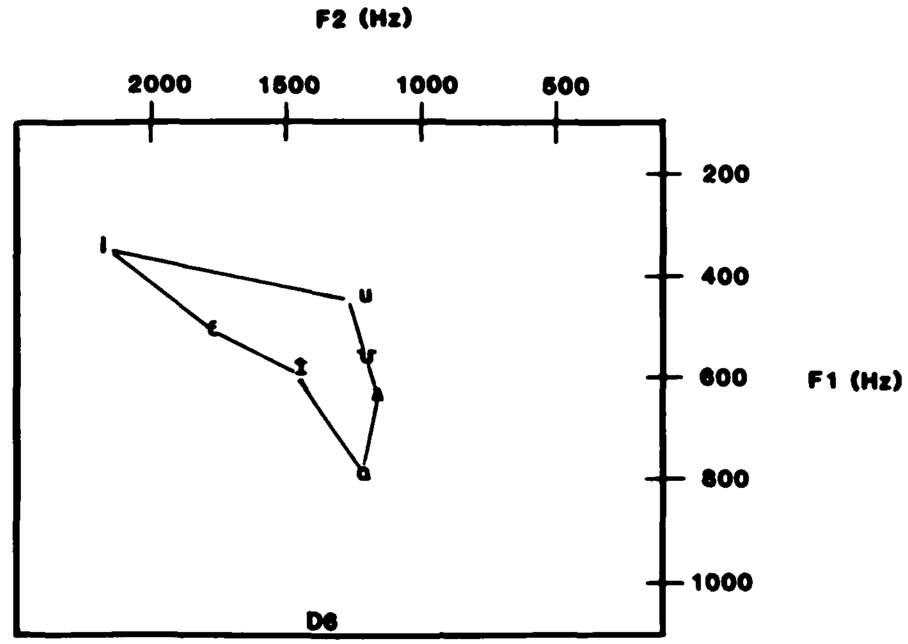


Figure 3. Average vowels for Talker D6.

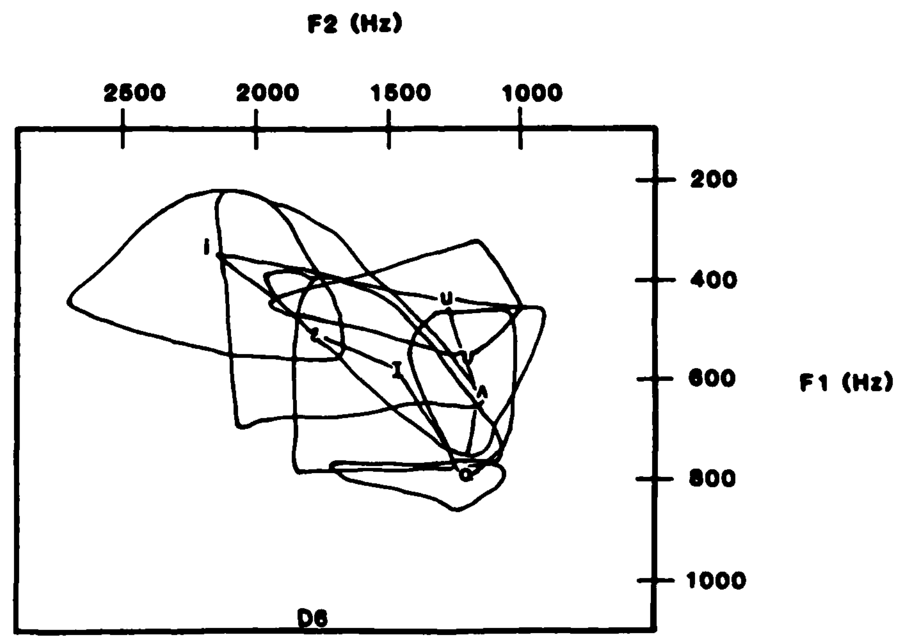


Figure 4. Range of vowels for Talker D6.

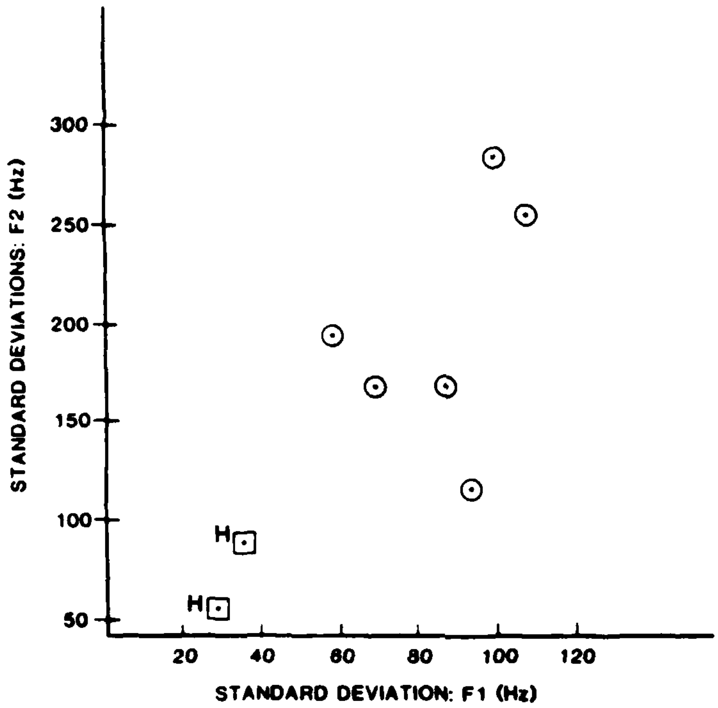


Figure 5. Standard deviations of F_1 vs. F_2 for all subjects.

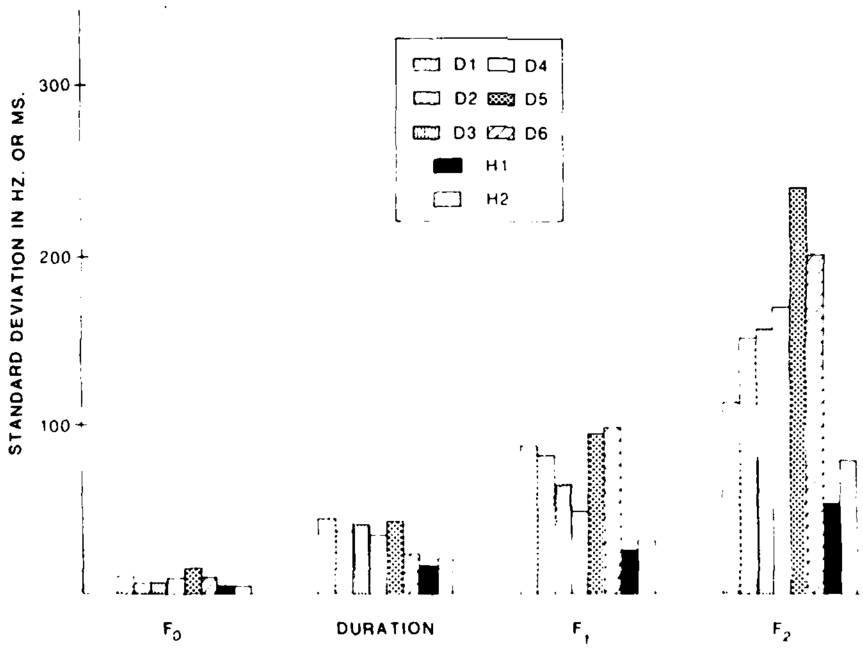


Figure 6. Standard deviations of F_0, duration, F_1 and F_2 for all subjects.

There remains the possibility that hearing-impaired talkers were using F_0, or duration, alone or in combination with F_1 and F_2 in their attempt to discriminate between vowels. This possibility was checked by comparing two linear discriminant analyses to see how many vowel targets can be discriminated using F_0 and duration, which were not discriminated by F_1 and F_2 alone. We find that for the most part, adding F_0 and duration information does not change the number of vowels that can be discriminated statistically, on a talker-by-talker basis. This provides additional support for Bush's (1981) finding that deaf talkers do not substitute F_0 differentiation for formant differentiation in vowel production.

Finally we turn to the perceptual part of the study. As we discussed above, a strong listener effect would be indirect evidence suggesting that deaf unintelligibility is due in part to a systematic, but deviant production strategy.

As Figure 7 shows, there was no statistically significant difference between experienced and inexperienced listeners. The listener effect for vowel identification has been reported by McGarr and Gelfer (1983), but not by Gulian and Hinds (1981). A listener effect for word identification has been found by Mangan (1961), Markides (1970), McGarr (1978), Nickerson (1973), and Thomas (1963).

Let us turn now to an examination of the effects of context. While the effects of context on vowel identification in normals has been the subject of debate in a voluminous literature (see, e.g., Ochiai & Fujimura, 1971; Pisoni, Carrell, & Simnick, 1979; Verbrugge, Strange, Shankweiler, & Edman, 1976), studies have at least suggested that phonetic context aids in recognition. That is the case here. Listeners, whether experienced or inexperienced, were most successful with sentences and syllables and least successful with gated segments excised from the vowel. Indeed the context effect is much more obvious for deaf than for hearing talkers.

Context also was important in the other judgment the listeners made, that is, whether the speaker was deaf or hearing. Since there were two hearing and six deaf speakers in the study, d' was used as a measure of the ability of listeners to identify the speakers as hearing or deaf, as shown in Figure 8. Again, the effects of experience were minimal. However, the listeners were increasingly correct in judging the speaker to be deaf as they had more dynamic information. This result qualitatively confirms Calvert's thesis result (1961). However, at a quantitative level, listeners in the present study could be shown to behave statistically slightly above chance levels in judging even isolated vowels. The ability of listeners to judge a vowel correctly was statistically independent of their ability to judge it as produced by a hearing or deaf child, whether the listener was experienced or inexperienced. This result again suggests that there is no special strategy that is effective in decoding deaf vowels.

Still another analysis was made of whether listeners were using conventional information in making vowel identity judgments for deaf talkers. Figures 9 and 10 show the acoustic data for the two individual deaf talkers discussed earlier, with circles around those vowels that are judged correctly at least 70% of the time. The effect of context is to enlarge the "correct vowel" area. Thus, we can speculate that placing a vowel within a consonant transition context allows the listener to be less dependent on precisely appropriate specification of vowel formant target information.

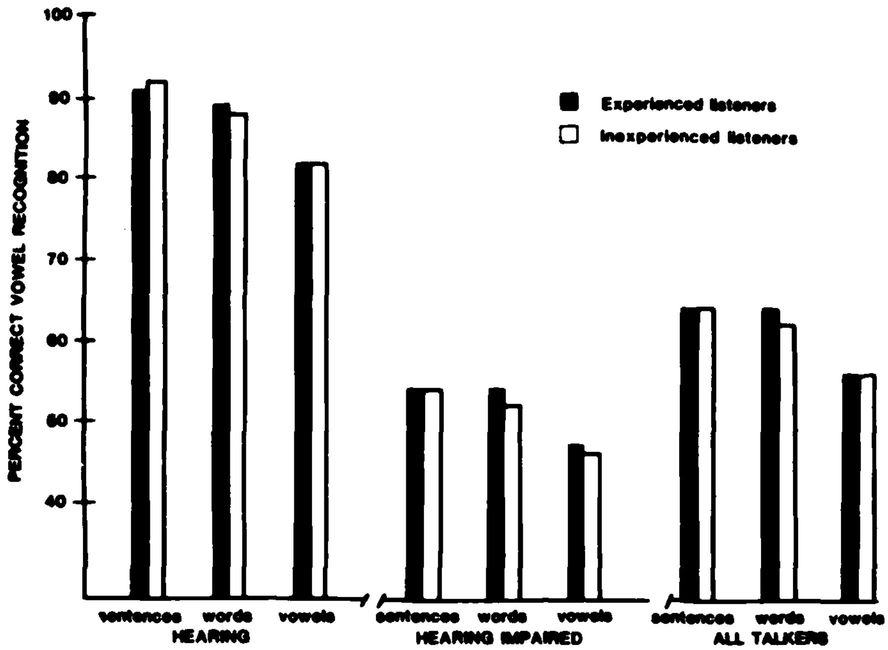


Figure 7. Effects of context on vowel recognition by experienced and inexperienced listeners, listening to hearing and hearing impaired talkers.

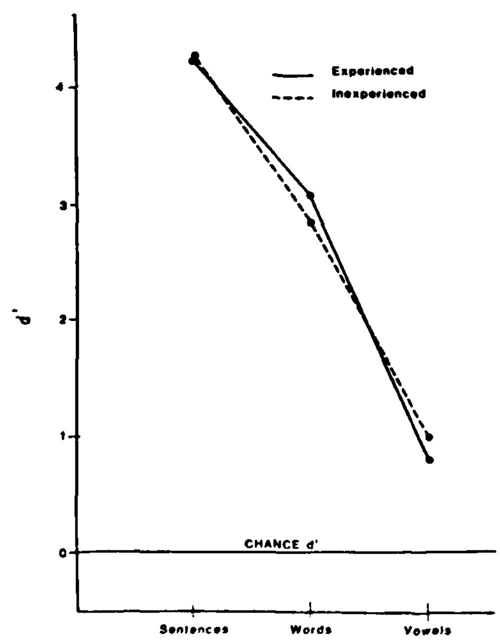


Figure 8. d' for vowels in various contexts.

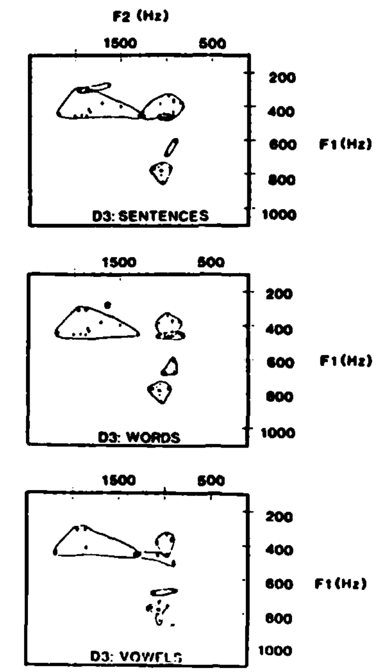


Figure 9. $F_1 \times F_2$ plots of vowel tokens perceived correctly in the three experimental contexts, for Speaker D3.

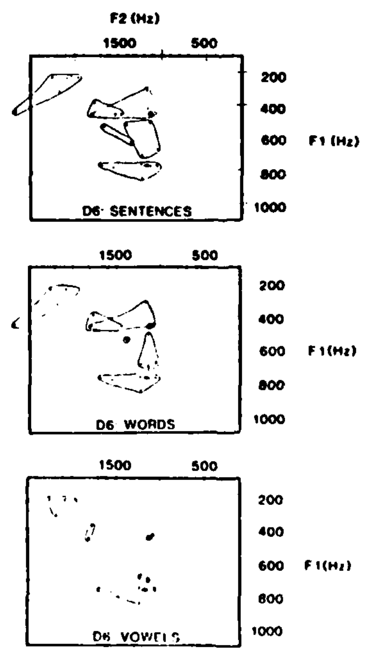


Figure 10. $F_1 \times F_2$ plots of vowel tokens perceived correctly in the three experimental contexts, for Speaker D6.

Let us summarize these results, and go on to say a bit about production. First, these analyses fail to provide any evidence that deaf speakers were using a substitution strategy in vowel production, or that experienced listeners were better than inexperienced, because of a different way of judging deaf speech. Deaf speakers were more variable than normals, although the pattern of variability was different from talker to talker. One interpretation of the results presented is that it is not appropriate to describe these talkers as presenting a deviant phonology. Indeed, we would argue that a "deviant phonology" description of their production does not capture essential aspects of their performance. The results we have seen for these children suggest that they are behaving, in a more extreme way, like normal children, as Kent (1976) describes them. Performance variability is an essential characteristic of all the speech of children as they learn to talk, and as they attain control of the production apparatus.

The nature of the articulator routines underlying the variability in acoustic output is unresolved by the study just described. However, we might note that the sequence of upper articulator movements in producing the utterance /bVb/ is fairly simple. The subject closes the lips for the initial and terminal bilabial consonants, and between these two gestures, s/he must produce an appropriate tongue configuration. If these gestures are produced in an inappropriately timed sequence, the acoustic result will be inappropriate, but the consequences of changing the relative timing of the gestural sequence is not directly represented in the acoustic signal.

One of the observations made by Ferguson and Farwell (1975) was that the attempts of a normal child to produce the word "pen" were variable precisely because she did not output the required sequence of articulatory gestures in the correct order. We believe that the characteristic variability in deaf speech may arise in part from the same sources (cf. McGarr & Gelfer, 1983; McGarr & Harris, 1983; McGarr & Löfqvist, 1982).

We illustrate this point with data from a tongue-lip coordination study of McGarr and Harris' (1983) in which stimuli not unlike Rubin's, (i.e., a bilabial-V-bilabial sequence) were used. Articulatory timing was monitored by electromyographic techniques. When muscle fibers contract, a change in potential is generated in the surrounding medium and these changes in potential can be measured by appropriately placed electrodes. Lip closure (e.g., in bilabial production) is accomplished in part by the contraction of the orbicularis oris muscle, a muscle whose fibers ring the lips. For production of a high vowel such as /i/, the tongue body is bunched and raised by contraction of the genioglossus, a muscle whose fibers radiate through the center of the tongue mass. The EMG record indicates this gesture sequencing.

Results for a hearing speaker producing the utterance /əpapip/ are shown in Figure 11. These data represent the ensemble average of about 20 repetitions or tokens of each utterance, with each token on the average showing essentially the same pattern of activity (see Harris & McGarr, 1980; McGarr & Harris, 1983). The line-up point, indicated by the vertical line at 0 ms, is the release burst of the second /p/. The data for the orbicularis oris (OO) show three well-defined peaks of activity corresponding to the lip gestures for the three /p/ closures in /əpapip/. The line-up point falls between the second and third peaks. For the genioglossus (GG), there is a peak of activity associated with /i/ but not /a/, because genioglossus is active in raising and bunching the tongue. Peak genioglossus activity occurs approximately at the acoustic line-up. This is not surprising because EMG

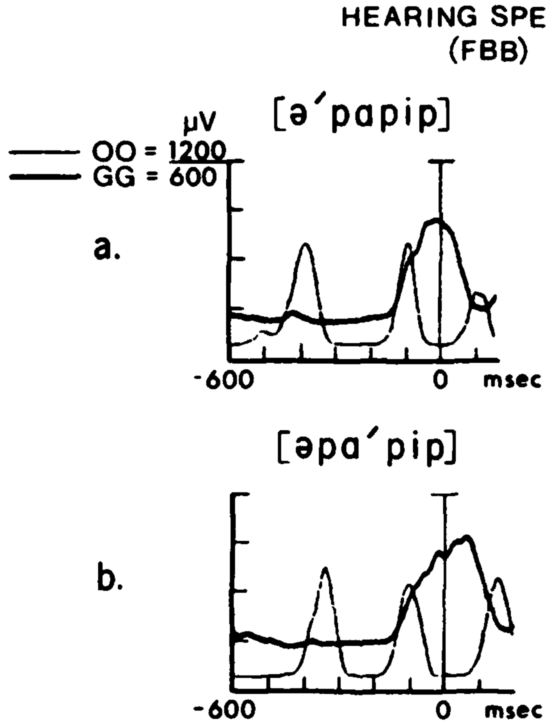


Figure 11. Average OO and GG outputs as a function of time, of simple nonsense utterances, for a normal talker. (Reproduced from Harris & McGarr, 1980.)

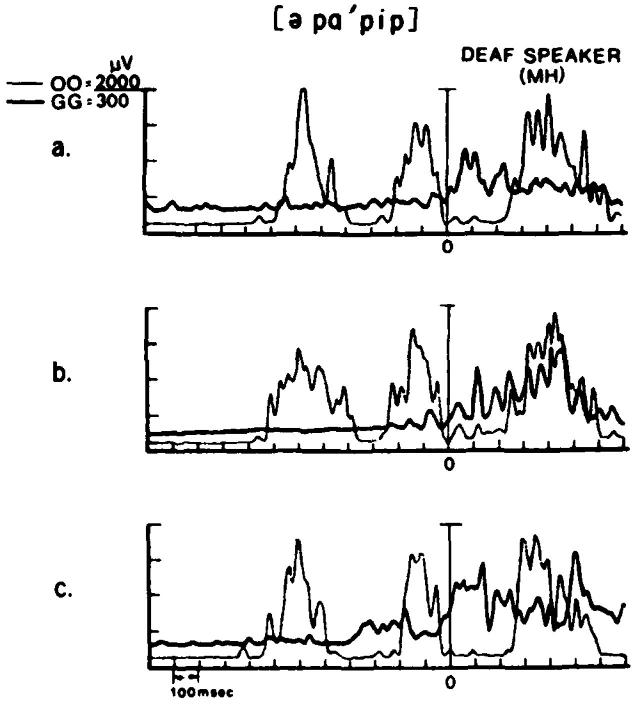


Figure 12. Three individual tokens of a simple nonsense utterance, showing OO and GG outputs as a function of time, for a hearing-impaired talker. (Reproduced from Harris & McGarr, 1980.)

activity typically precedes the articulatory event to which it is attached by about 50 to 100 ms. Shifting stress from the first (Figure 11A) to the second vowel (Figure 11B) does not disrupt this temporal relationship.

Figure 12 shows similar data for an oral deaf adult. The EMG pattern for OO shows, as for the hearing subject, three well-defined peaks of activity. The duration of the peaks is prolonged, however. In Figure 12A, peak GG activity occurs between the second and third orbicularis oris peaks but is late relative to the acoustic event. This pattern was most like normal. In Figure 12B the GG activity was too late. In Figure 12C, activity begins during what should be /a/ production, when the GG should be silent. Thus, the EMG pattern for GG is quite variable from token to token. This variability is reflected in a less well-defined average pattern (see McGarr & Harris, 1983, for more details).

While this evidence is fragmentary, it suggests precisely the sort of production variability we might expect; that is, while the behavior of a visible articulator is more or less normal, activity for one of the muscles associated with tongue movement is variable in its temporal alignment with the activity of the visible articulator. This could produce the kind of acoustic variability analyzed in Rubin's work. Similar interarticulator variability has also been described in our work with deaf speakers for larynx-upper articulators (McGarr & Löfqvist, 1982) and tongue-lip (McGarr & Gelfer, 1983) coordination.

One final result illustrates the extraordinary stability of interarticulator timing in normal adult speech production. Tuller (Harris, Tuller, & Kelso, 1985; Tuller & Kelso, 1984; Tuller, Kelso, & Harris, 1982, 1983) has performed a series of experiments in which normal adult subjects produce simple nonsense syllables (again, of the form /papap/), with stress on either the first or second syllable and at two self-selected speaking rates. In a typical experiment, lip and jaw movements were monitored by fixing light-emitting diodes on these articulators. In a utterance such as /babab/, downward jaw movements can be associated with vowels, while upward lip movement can be associated with consonants. Tuller was thus able to examine the relationship of the temporal onset of the medial consonant to the duration of a vowel-to-vowel interval.

Figure 13 shows the data plots with the values of $\underline{r}$ and the slopes for a linear regression for four utterance types, /bapab/, /babab/, /bawab/, and /bavab/ for a single speaker. The $\underline{r}$ values do not vary systematically with consonant. For the various measures analyzed, the Pearson product-moment correlation values range from +.84 to +.97 across the four subjects of the experiment. While the values of $\underline{m}$ show a trend towards flatter slopes and thus earlier consonant onsets for /v/ and /w/ as compared to /p/ and /b/, the ordering of slopes was not identical across subjects.

The substantial size of the linear correlations suggests that stability of the ratio over changes in vowel duration produced by stress and speaking rate changes is a characteristic of mature normal speech production. If we were to examine similar data for normal children, we would expect a systematic decrease in the scatter around the line of best fit with increasing articulatory maturity. For deaf speakers, we would expect even lower correlation values. To substantiate this, we are presently analyzing data from a comparative study of deaf and normal speakers.

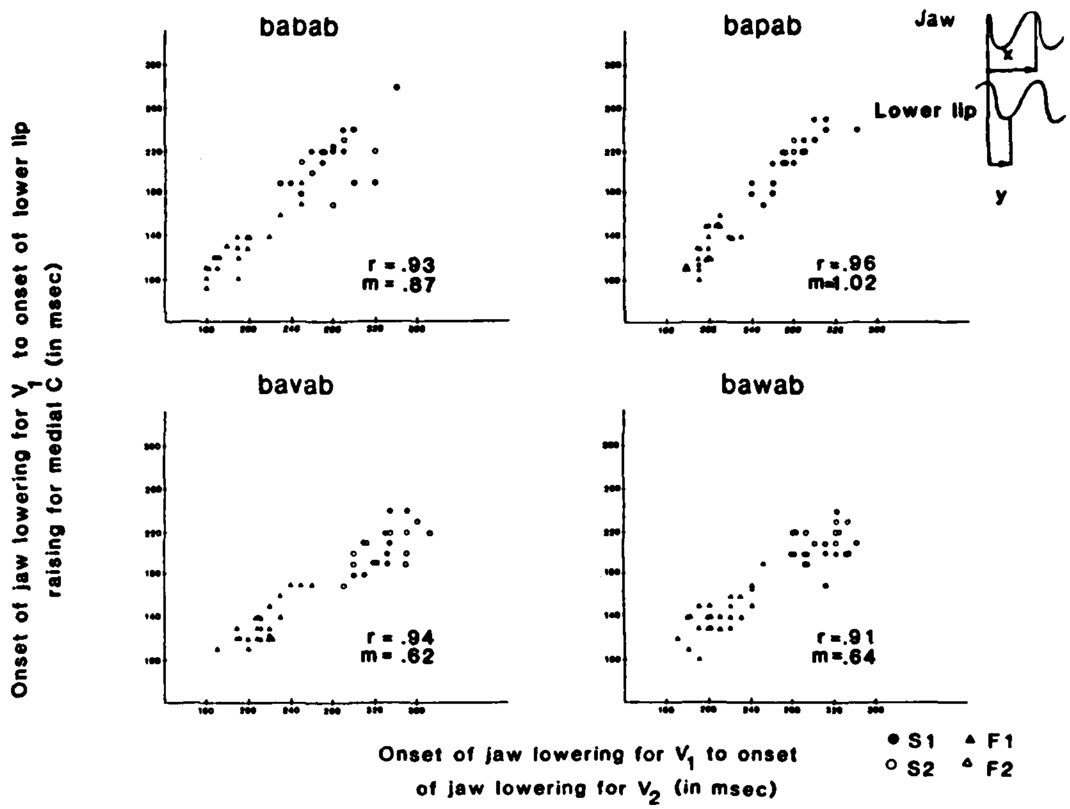


Figure 13. Period (jaw lowering) versus latency (lower lip raising) for nonsense disyllables differing in medial consonant for a single subject. Circles indicate utterances spoken at a conversational rate, triangles indicate a somewhat faster rate. Filled symbols have stress on the first syllable, open symbols have stress on the second syllable (Data for Subject EH described in Tuller and Kelso, 1984).

Finally let us return to the beginning of this paper and point to the moral. Although "deaf speech" may have distinctive characteristics, the striking thing about the results reported here is the link between deaf speech and motorically immature speech. This relationship will in part be obscured by any description that ignores variability as an essential characteristic of the speech production capabilities.

References

Angelocci, A., Kopp, G., & Holbrook, A. (1964). The vowel formants of deaf and normal hearing eleven to fourteen year old boys. Journal of Speech and Hearing Disorders, 29, 156-170.

Bush, M. (1981). Vowel articulation and laryngeal control in the speech of the deaf. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Calvert, D. R. (1961). Some acoustic characteristics of the speech of profoundly deaf individuals. Unpublished doctoral dissertation, Stanford University.

Edwards, M. L., & Shriberg, L. D. (1983). Phonology: Applications in communicative disorders. San Diego: College-Hill Press.

Eguchi, S., & Hirsh, I. J. (1969). Development of speech sounds in children. Acta Otolaryngologica Supplement 25 .

Eimas, P. D. (1982). Speech perception: A view of the initial state and perceptual mechanics. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), Perspectives on mental representation (pp. 339-360). Hillsdale, NJ: Erlbaum.

Elbert, M., Dinnsen, D. A., & Weismer, G. (1984). Phonological theory and the misarticulating child. ASHA Monograph 22.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. Language, 51, 419-439.

Fisher, J., King, A., Parker, A., & Wright, R. (1983). Assessment of speech production and speech perception as a basis for therapy. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training and personnel preparation (pp. 195-214). Baltimore, MD: University Park Press.

Gulian, A. B., & Hinds, J. (1981). Response bias in the perception of deaf speech by experienced and naive listeners. British Journal of Audiology, 15, 165-171.

Harris, K. S. (1983). A review of Phonological interventions: Concepts and procedures. Applied Psycholinguistics, 4, 294-297.

Harris, K. S., & McGarr, N. S. (1980). Relationships between speech perception and speech production in normal hearing and hearing impaired subjects: Implications for training speech of the hearing impaired. In J. D. Subtelny (Ed.), Speech assessment and speech improvement for the hearing impaired (pp. 316-337). Washington, DC: A. G. Bell Association.

Harris, K. S., Tuller, B., & Kelso, J. A. S. (1985). Temporal invariance in speech production. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum.

Ingram, D. (1976). Phonological disability in children. New York: Elsevier.

Jakobson, R. (1968). Child language, aphasia, and phonological universals. The Hague: Mouton.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. Journal of Speech and Hearing Research, 19, 421-447.

Mangan, K. (1961). Speech improvement through articulation testing. American Annals of the Deaf, 106, 391-396.

Markides, A. (1970). The speech of deaf and partially hearing children with special reference to factors affecting intelligibility. British Journal of Disorders of Communication, 5, 126-140.

Maxwell, E. M. (1979). Competing analyses of deviant phonology. Glossa, 13, 181-214.

McGarr, N. S. (1978). The differences between experienced and inexperienced listeners in understanding the speech of the deaf. Unpublished doctoral dissertation, The City University of New York.

McGarr, N. S., & Gelfer, C. (1983). Simultaneous measurements of vowels produced by a hearing-impaired speaker. Language and Speech, 26, 233-246.

McGarr, N. S., & Harris, K. S. (1983). Articulatory control in a deaf speaker. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training, and personnel preparation. Baltimore: University Park Press.

McGarr, N. S., & Löfqvist, A. (1982). Obstruent production in hearing-impaired speakers: Interarticulator timing and acoustics. Journal of the Acoustical Society of America, 72, 34-42.

Menn, L. (1980). Phonological theory and child phonology. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson, (Eds.), Child phonology Vol. 1 Production (pp. 23-41). New York: Academic Press.

Menyuk, P., & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), Language acquisition (pp. 49-70). Cambridge: Cambridge University Press.

Monsen, R. B. (1976). Normal and reduced-phonological space: The production of English vowels by deaf adolescents. Journal of Phonetics, 4, 189-198.

Monsen, R. B. (1983). General effects of deafness on phonation and articulation. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training and personnel preparation. Baltimore, MD: University Park Press.

Monsen, R., & Engebretson, M. (1983). The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear production. Journal of Speech and Hearing Research, 26, 89-97.

Nickerson, R. (1973). Computerized speech training aids for the deaf. Bolt, Beranek and Newman, Inc. Report No. 2633.

Ochiai, K., & Fujimura, O. (1971). Vowel identification and phonetic contexts. Reports of the University of Electro-Communication 22, 103-111.

Pisoni, D., Carrell, T. D., & Simnick, S. S. (1979). Does a listener need to recover the dynamic vocal tract gestures of a talker to recognize his vowels? Journal of the Acoustical Society of America, 22, 807-820.

Rubin, J. (1983). Static and dynamic information in vowels produced by the hearing impaired. Unpublished doctoral dissertation, The City University of New York.

Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. Journal of Phonetics, 6, 37-67.

Smith, N. V. (1973). The acquisition of phonology: A case study. London: Cambridge University Press.

Stevens, K. N., Nickerson, R. S., & Rollins, A. M. (1983). Suprasegmental and postural aspects of speech production and their effect on articulatory skills and intelligibility. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training and personnel preparation. Baltimore, MD: University Park Press.

Studdert-Kennedy, M. (1985). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum.

Thomas, W. (1963). Intelligibility of the speech of deaf children. Proceedings of the International Congress on the Education of the Deaf. Document No. 196.

Tuller, B., Kelso, J. A. S., & Harris, K. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1983). Further evidence for the role of relative timing in speech. Journal of Experimental Psychology: Human Perception and Performance, 9, 829-833.

Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariance. Journal of the Acoustical Society of America, 76, 1030-1036.

Verbrugge, R. R., Strange, W., Shankweiler, D. P., & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? Journal of the Acoustical Society of America, 60, 198-212.

CAN LINGUISTIC BOUNDARIES CHANGE THE EFFECTIVENESS OF SILENCE AS A PHONETIC CUE?*

Bruno H. Repp

Abstract. This study investigated the influence of three kinds of linguistic boundaries--word boundaries, prosodic breaks, and syntactic breaks--on the perception of a silent interval at the boundary site as a cue to the presence of a labial stop consonant. The experimental technique involved cross-splicing portions of four naturally produced pairs of sentences, as well as presentation of excerpts from these sentences. Although one sentence pair showed a pronounced syntactic boundary effect, the other three (including two that were better controlled for semantic bias) did not, which points to a different, stimulus-specific origin of the effect obtained. Prosodic boundary effects were also generally absent, presumably because the stimuli were constructed such that prosodic variation ceased 78 ms prior to the critical silent interval. Only introduction of a word boundary effected a systematic reduction in stop consonant percepts, although this manipulation was confounded with other contextual factors. On the whole, the data provide little evidence for any direct effects of structural linguistic variables on phonetic segment perception; such effects seem to be restricted to the level of word recognition.

1. Introduction

One fundamental issue in speech perception research concerns the relative importance of physical signal properties ("bottom-up" information) versus the listener's expectations and interpretations ("top-down" processes). There is little doubt that phonotactic, semantic, and pragmatic factors can influence word perception, particularly when the speech signal is ambiguous (see, e.g., Fox, 1984; Ganong, 1980; Massaro & Cohen, 1983). Whenever a listener has internally generated or contextually induced expectations about the likelihood of certain phonological or lexical alternatives, these expectancies will help reduce any uncertainty introduced by insufficient physical information.

It is much less clear whether a listener's apprehension of structural factors that do not affect the likelihood of phonological or lexical alternatives can have repercussions at the level of phonetic segment perception. Specifically, the question is whether linguistic boundaries (syllabic, lexical, or syntactic) can reduce the phonetic coherence of an utterance at the boundary site, with possible consequences for the perceived segmental composition. Such an interaction, if it were to occur, would be theoretically interesting, for it would suggest that higher-level processes of

---

*Journal of Phonetics, in press.

Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR-05596 to Haskins Laboratories. I am grateful to Andrea Levitt and Patti Price for critical comments.

lexical access and syntactic analysis can exert a direct influence on the internal representation of the bottom-up information, or at least generate expectations about its detailed acoustic structure. It should be kept in mind, however, that the effects under investigation, unlike the top-down effects studied extensively in research on word recognition (see, e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980), are rather special phenomena that, even if real, probably play only a very minor role in real speech understanding.

The evidence for such effects, however, is not compelling so far. Previous studies on this topic have been concerned with the function of silence as a phonetic cue. There is much evidence that short periods of silence in speech are not perceived as gaps or interruptions but as carriers of articulatory information about closure of the vocal tract, as occurs in connection with stop and affricate consonants (see, e.g., Dorman, Raphael, & Liberman, 1979; Repp, Liberman, Eccardt, & Pesetsky, 1978). One particular situation investigated in several recent studies involves the effect of a short interval of silence preceding a fricative noise as a cue to the contrast between a word-initial fricative and affricate (Dechovitz, 1979, 1980, 1981; Price & Levitt, 1983; Rakerd, Dechovitz, & Verbrugge, 1982). The hypothesis tested in these studies was that introduction of a coincident linguistic break might reduce the perceptual effectiveness of the silence, either because the silence could be interpreted as a hesitation associated with the break rather than as an articulatory closure associated with a stop consonant, or because the linguistic boundary has a direct disruptive influence on the coherence of the signal portions preceding and following the silence, so that the presence and precise duration of the closure interval become perceptually irrelevant. Dechovitz (1979, 1980, 1981) claimed to have found such an effect due to syntactic structure alone--i.e., he found a significant reduction of silence-cued affricate percepts when a syntactic boundary at the critical location was created by remote context under semantically neutral and constant local acoustic conditions. These data have not been published, however, and Price and Levitt (1983) have failed to replicate the effect.

All these previous studies, however, found that the introduction of clause- or sentence-final prosody--including a falling intonation contour and final syllable lengthening--reduced the perceptual effect of a following silent interval. Although prosodic changes usually accompany changes in linguistic structure and thus carry considerable lexical and syntactic information, they do involve acoustic changes in the immediate vicinity of the silent interval. Since this may alter some of the local phonetic cues, the observed prosodic effects may not represent an influence of perceived linguistic structure on phonetic perception but may have more direct causes.

The present experiment extended these earlier studies by further investigating the influence on phonetic perception of syntactic and prosodic breaks, and by also considering the possible role of word boundaries. Stimulus materials were chosen in which the critical silence served as a cue for a labial stop consonant following a fricative and preceding a liquid (see Fitch, Halwes, Erickson, & Liberman, 1980; Dorman et al., 1979). The fricative-affricate contrast used previously is characterized by a rather sharp category boundary at a very short silence duration, which raises the possibility of psychoacoustic interactions that are immune to contextual influences. The type of contrast employed here, on the other hand, typically has its category boundary at relatively longer silence durations, so low-level psychoacoustic interactions are unlikely (see Pastore, Szczesiul, & Rosenblum,

1984; Repp, 1985), and it also has a larger region of ambiguity, which makes it more sensitive to influences of all kinds.[1] The critical silence was embedded in plausible, natural sentences, which constitutes an improvement over the somewhat contrived and limited materials used in earlier investigations. Syntactic and prosodic factors were varied independently by swapping the two words surrounding the critical silent interval between syntactically different sentence frames. (Prosodic variation beyond these two words was confounded with syntactic structure.) Prosodic variation in the word immediately preceding the silence included the duration and amplitude envelope of the final [s] noise segment, which--judging from earlier findings and from informal observations during stimulus construction--would certainly have had a strong perceptual effect. Because of this foregone conclusion, it was decided to neutralize this segment and to examine only whether prosodic information <u>beyond</u> the immediately preceding acoustic segment can influence phonetic perception.

It should be pointed out that the role of silence as a cue to stop consonant perception is twofold. If the closure silence is too short (less than about 60 ms in the fricative-liquid context), no stop consonant may be perceived even when other cues are available (e.g., Dorman et al., 1979; Fitch et al., 1980). If the silence is longer (roughly 100-300 ms), a (labial) stop consonant will often be perceived even when there are no other cues (Dorman et al., 1979; Repp, 1985). These two effects may be called "stop suppression" and "stop generation," respectively (Repp, 1985). The stop suppression effect may in part be due to psychoacoustic interactions (such as forward masking) between the closely adjacent signal portions (however, see Pastore et al., 1984), whereas such interactions are much less likely in the case of the stop generation effect. Therefore, if there are any effects of linguistic boundaries on phonetic perception, they are more likely to occur at longer closure intervals, where psychoacoustic interactions play no role. The specific hypothesis tested was that, compared to a no-boundary condition, introduction of a linguistic boundary at the point of the critical silence would <u>decrease</u> the number of stop consonant responses at relatively long closure durations. To the extent that stop suppression is <u>not</u> caused by psychoacoustic interactions, an <u>increase</u> of stop responses might be predicted at short closure durations, because linguistic boundary might then reduce the (negative) cue value of short silences as well.

Following some piloting, two full-size experiments were conducted that were very similar in design. Because stimulus parameters were still not optimal, the first experiment inadvertently focused exclusively on the region of stop consonant suppression, where little sensitivity to linguistic boundaries was expected (and obtained). Therefore, only the results of the second experiment will be reported, which--due to additional stimulus adjustments--successfully encompassed both regions of stop consonant generation and suppression. Where the two designs overlapped, the results of the first experiment were consistent with the findings reported below.

2. Methods

2.1. Subjects

Ten paid volunteers participated. All were Yale undergraduates and native speakers of American English.

2.2. Stimulus Preparation

The stimulus sentences are shown in Table 1. Four pairs of sentences were constructed. The members of each pair contained the same two critical words in succession; the first word ended in [s], whereas the second word either did or did not begin with [b], so that there were two versions of each sentence. In one sentence of each pair (version b), a clause boundary intervened between the two critical words, whereas in the other sentence (version a), the two words formed a syntactic unit. The second critical word, which either did or did not begin with [b], represented fictitious surnames in two instances (Nos. 2 and 3) and real words in the other two (1 and 4). Orthogonal to this distinction, the consonant following the optional [b] was [l] in two words (1 and 3) and [r] in the other two (2 and 4). Because of the two possible versions of the second critical word, there was a total of 16 sentences.

Table 1

Stimulus Sentences

1. a. The royal tomb was protected by six (b)locks of solid gold.
   b. When the clock strikes six, (b)lock the gate.

2. a. The girl tried to kiss (B)Radford on the cheek.
   b. After giving his wife a kiss, (B)Radford boarded the train.

3. a. Will you please welcome Miss (B)Lackman to the office.
   b. Enraged by a spectacular miss, (B)Lackman quit the game.

4. a. To the maid's dismay, worse (b)rooms could hardly be imagined.
   b. What made matters worse, (b)rooms were difficult to find.

These 16 sentences were recorded by a male speaker of American English in a sound-insulated booth using high-quality equipment. The recordings were low-pass filtered at 4.9 kHz and digitized at a 10 kHz sampling rate. Using a waveform editor in conjunction with careful listening, each sentence was divided into four sections that were stored in separate computer files: preceding context (C1), first critical word (W1), second critical word (W2), and following context (C2). All cuts were made at zero crossings. In those sentences in which W2 had an initial [b], the stop closure was edited out and discarded. Thus, W1 ended at the beginning of the stop closure and W2 began at its end. In sentences without a W2-initial [b], the end of W1 and the beginning of W2 coincided, except in two sentences in which a lateral noise burst occurring at an [s-l] juncture was edited out.

For each sentence pair listed in Table 1, each of the two different context frames (C1+C2) existed in two distinct productions. Only one of these was retained--that deriving from sentences in which W2 had been articulated with an initial [b] (an arbitrary choice). The first critical word (W1) existed in four recorded versions; only those two versions that were not followed by a W2-initial [b] were used (another arbitrary choice). In these

two remaining versions, the clause-final [s] noises were much longer in duration (ranging from 144 to 201 ms across the four sentences) than the non-clause-final [s] noises (range: 55 to 109 ms). For reasons outlined in the Introduction, these noises were removed and replaced by a constant [s] noise excerpted from the same talker's production of the word "spectacular" (sentence 3b). (For an explanation of this choice, see below and footnote 2.) This [s] noise, originally only 54 ms in duration, was artificially lengthened to 78 ms by duplicating a 24-ms central section of the waveform.

Finally, the onsets of the W2 words, which existed in four recorded versions, were examined and edited. Words articulated with an initial [b] all had labial release bursts ranging in duration from 12 to 18 ms. These bursts, which provided strong stop manner cues (see, e.g., Repp, 1984a) were eliminated, leaving only potential coarticulatory cues in the periodic stimulus portion. The words without an initial [b] had no bursts and were retained without change.

In summary, then, for each of the four sentence pairs listed in Table 1, there were two different context frames C1+C2, each in a single recorded version; two versions of W1, a clause-final one and a non-clause-final one, with a common final [s] noise; and four versions of W2, two that had originally started with [b] and two that had not, and orthogonal to this distinction, two clause-initial and two non-clause-initial ones.

These components were re-assembled into sentences, with four different silent closure intervals introduced between the W1 and W2 words: 40, 80, 120, and 160 ms. All possible combinations of sentence components were employed in the <u>sentence</u> test, leading to a total of 4 (sentence types) X 2 (contexts) X 2 (W1) X 4 (W2) X 4 (silences) = 256 sentences. They were recorded in 4 blocks of 64, randomized within each block in groups of 16, with interstimulus intervals (ISIs) of 3 s and intervals of 10 s between groups. The first and third blocks contained sentences in which the prosody of W1 was appropriate for the syntactic context, whereas the second and fourth blocks contained the sentences in which W1 had the inappropriate prosody. These latter sentences sounded somewhat odd but not bizarre; they were deemed appropriate for an assessment of prosodic factors.

In addition to this lengthy main test, four shorter test tapes were recorded. The first of these was a <u>pretest</u> containing 16 sentences. The first 8 sentences represented the eight different contexts, with prosodically appropriate W1 and W2; W2 was either the "stronger" version (i.e., that originally began with [b]) preceded by the second-shortest silence (80 ms), or the "weaker" version (that originally began with [l] or [r]) preceded by the longest silence (160 ms). The second set of 8 sentences contained the context-W2 combinations not contained in the first set. All 16 sentences were arranged in a quasi-counterbalanced sequence, with ISIs of 20 s. The purpose of this pretest was to assess the listeners' response to the test sentences on first hearing.

The second test contained the W1-silence-W2 <u>word pairs</u> in all 128 possible combinations, without their sentential context. They were recorded in 4 blocks of 32, with ISIs of 4 s. The purpose of this test was to provide a baseline for assessing the contribution of the contextual frame, regardless of its syntactic implications, and to examine prosodic effects in this more restricted context (cf. Price & Levitt, 1983).

In the third test, the W2 words were preceded only by the constant [s] noise plus silence, to provide a baseline for testing the hypothesis that a word boundary following the [s] reduces the likelihood of silenced-cued labial stop percepts. In this test, the [s] was to be perceived as the initial segment of a nonsense word (e.g., "splock"). The constant [s] noise was taken from a word-initial position (see above) to facilitate this task.[2]

Finally, the excerpted W2 words were assembled into a single-word test. The 16 W2 words (4 words X 4 versions) were recorded in 4 different random orders with ISIs of 4 s. This test was to provide a baseline against which the effect of closure silence in the other tests could be compared.

2.3. Procedure

The subjects listened to all tests in a single session, using TDH-39 earphones in a quiet room. The tests were presented in a fixed sequence: The pretest was followed by the sentence test, the word pair test, the nonsense word test, and the single word test.

In the pretest, the subjects' task was to write each sentence down verbatim on a blank sheet of paper. Subjects were informed that the sentences were meaningful, that some of them contained proper names, and that the second set of 8 would be very similar--but not necessarily identical--to the first set of 8.

For the sentence test, the subjects were provided with printed answer sheets. Each page listed all the stimulus sentences on top, arranged as in Table 1, without the italics but with two words in each sentence capitalized. The first of those words was a key word in the first clause (e.g., ROYAL) identifying the context; the second was W2. For each item the answer sheets listed the four pairs of possible key words and W2 below each pair, with the initial B in parentheses. The subjects' task was, for each sentence heard, first to circle the appropriate key word and then to indicate, by either circling or crossing out the parenthetical B in the word below, whether W2 did or did not begin with a [b]. Since the sentences came at a fairly brisk rate, the subjects were encouraged to circle the key word before the sentence was over, and to skip the key word if the time seemed too short. Some subjects omitted a few key word responses in the beginning but soon found their rhythm. The only purpose of the key word responses was to keep the subjects' attention on the context and thus to prevent an overly selective listening strategy.

For the word pair test, answer sheets listed for each item the four possible W1-W2 pairs, with the W2-initial B in parentheses. The subjects' task was to find the appropriate word pair and either to circle or cross out the B. For the nonsense word test, the answer sheet listed for each item the four possible choices with a parenthetical P following the initial S (i.e., S(P)LOCK, S(P)RADFORD, S(P)LACKMAN, S(P)ROOMS). Subjects were asked to try their best to consider the stimuli as [s]-initiated nonsense words and to either circle or cross out the P in the correct alternative. Their attention was drawn to the unfamiliar [sr] cluster as a possible beginning of a nonsense word. Finally, the answer sheet for the single word test listed the four possible W2 choices for each item, and subjects located the correct alternative and either circled or crossed out the parenthetical word-initial B.

3. Results and Discussion

3.1. General Contextual Effects

Averaging over different versions of W1 and W2, Figure 1 shows the results in terms of percent labial stop responses to W2 onset, separately for each sentence pair (S1-S4), as a function of silent closure duration. The various response functions compare sentences with (S-b) and without (S-a) a syntactic break preceding W2, word pairs (WP), and nonsense words (NW). The percentage of "b" responses to single W2 words (SW) is indicated by the arrows at the right-hand side of each panel.

The first thing to note is that the percentage of labial stop percepts increased as closure duration increased. Repeated-measures analyses of variance on the separate tests showed that this expected effect was extremely significant and also interacted strongly with the Sentence factor, as is evident from the different slopes of the response functions (all effects at least $p < .001$). A visual comparison with the single-word (SW) percentages shows that labial stop responses at the longer closure durations exceeded those to single W2 words by a considerable margin (the stop generation effect), whereas the opposite relationship held at the shortest silent interval (the stop suppression effect).

The next finding to note in Figure 1 is that the response functions for word pairs (WP) were not systematically different from those for sentences (S-a and S-b combined); thus, having some sentential context around the W1-silence-W2 constellation did not influence the subjects' criterion for reporting a "b." By contrast, the percentages of labial stop responses were much higher in [s]-initiated nonwords (NW) than in the other conditions, where a word boundary separated the [s] from the following context. (The exception is Sentence 3, where a ceiling effect may have prevented a difference from emerging.) In a combined analysis of the WP and NW conditions, the main effect of Condition was highly significant, $F(1,9) = 51.18$, $p < .0001$, and so were its interactions with Sentences, Closure Duration, and both of these factors (all $p < .0004$ or less, mainly due to the different pattern for sentence 3). The interaction with Closure Duration reflected the fact that the effect was smallest at the shortest silence duration; there was no tendency toward a reversed effect in the stop suppression region, which suggests some psychoacoustic limit at short silences. A response bias against the unfamiliar "sr" clusters in nonwords could have operated in sentences 2 and 4, but not in sentence 1. Thus, unless the immediate context preceding the [s] (i.e., W1) had some direct influence on subjects' criteria, apart from introducing a word boundary, these results may be interpreted as supporting the hypothesis that the linguistic factor of word juncture attenuated the cue value of longer silences as a positive stop manner cue.

3.2. Syntactic Effects

Turning now to the comparison of syntactic conditions, it is evident from Figure 1 that there was a large and consistent difference between the two versions of sentence 1, with the syntactic boundary version (S-b) receiving fewer "b" responses. However, none of the other three sentences showed such a consistent difference. This pattern of results was reflected in a highly significant Sentence X Context interaction, $F(3,27) = 10.7$, $p < .0001$, whereas the main effect of Context was not significant. Separate analyses of variance for individual sentences showed a significant effect of Context for sentence

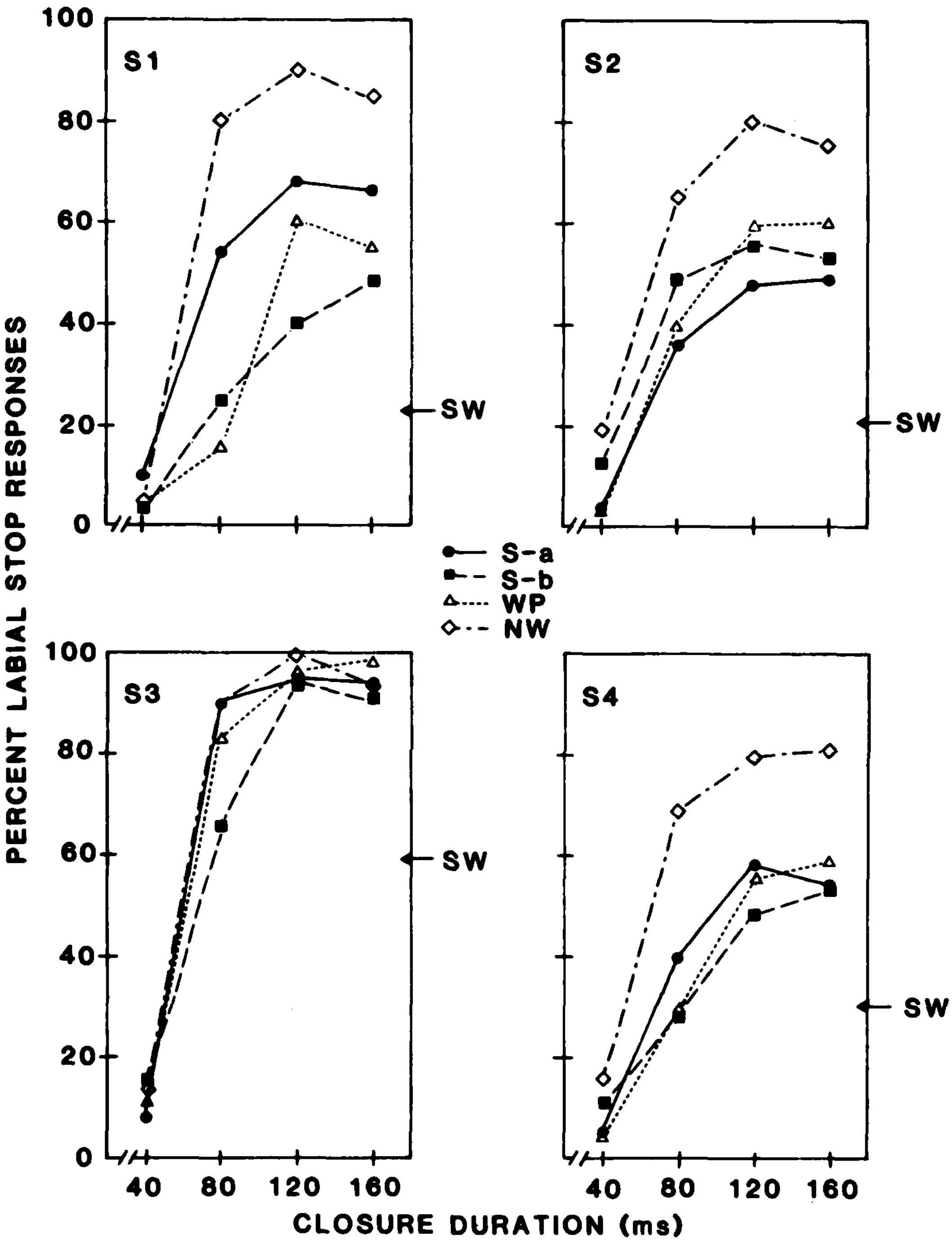


Figure 1. Percent stop responses, separately for the four sentences (S1-S4), as a function of closure duration. Separate response functions are shown for sentences without a syntactic boundary (S-a), with a syntactic boundary (S-b), for isolated word pairs (WP), and for nonsense words (NW). The arrows represent the percentages for single words (SW).

1, $F(1,9) = 11.51$, $p < .01$, but no significant effects for any of the other sentences.[3] Since sentences 2 and 3, because of the use of proper names as W2, were semantically better controlled than sentences 1 and 4, these results do not support the hypothesis of a syntactic influence on phonetic perception. Rather, they suggest that there was something peculiar about sentence 1.

The most likely possibility is that the two alternatives of W2 were not equally plausible in the two semantic contexts of sentence 1, "six blocks of gold" being more acceptable than "six locks of gold," and perhaps also "lock the gate" being preferred to "block the gate." This possibility was assessed by presenting versions a and b of sentences 1 and 4 in written form to 20 staff members of Haskins Laboratories, with the request to choose the W2 alternative that "fits better into the sentence frame (i.e, that makes the sentence more meaningful, more likely, or more appealing)." To counteract order effects, two versions of this short test were used, with reversed orderings of the sentences and of the W2 alternatives for each sentence. The results revealed that "block(s)" indeed was considered relatively more plausible in sentence 1a (8 out of 20 responses) than in sentence 1b (0 responses). A similar asymmetry was obtained for sentence 4: "brooms" was preferred in sentence 4a (10 responses) relative to sentence 4b (1 response). Although sentence 4 did not show a significant "syntactic" effect in the sentence test, there was a tendency in that direction (Figure 1). Therefore, the "syntactic" effect in sentence 1 is attributed very tentatively to a semantically conditioned response bias.[4]

3.3. Prosodic Effects

The absence of consistent syntactic effects implies also that prosodic variation in the sentence frame preceding W1 had no systematic effect (with the possible exception of sentence 1). In addition, however, it was quite obvious from the data that W1 prosody itself had very little effect. The effect of appropriate vs. inappropriate prosody (with respect to syntactic structure) should have been revealed in a Context X W1 interaction in the sentence test. This interaction was nonsignificant. There could also have been an effect due to W1 intonation per se (clause-final vs. non-clause-final), regardless of its context appropriateness. The W1 main effect, however, was likewise nonsignificant in both the sentence and word pair tests. Moreover, no individual sentence showed any pronounced prosodic effect. This was surprising, since earlier studies (Dechovitz, 1979; Rakerd et al., 1982) had found strong prosodic effects, and the present technique of cross-splicing might have been expected to introduce artifactually large effects.[5]

One important possibility to consider is that clause-final and non-clause-final versions of W1 simply did not differ much, apart from the original difference in final [s] duration (see Methods section), which had been neutralized. To examine this issue, temporal measurements were obtained from the W1 waveforms and are shown in Table 2. It is clearly evident that clause-final versions (b) of W1 had substantially longer durations and lower terminal fundamental frequencies than non-clause-final versions (a). The durational differences extended over all acoustic segments of the W1 syllable, including of course the final [s] prior to its neutralization (not shown in Table 2). Thus there was a clear basis for potential prosodic effects due to W1.

Table 2

W1 Durations (Not Including the Final [s] Noise) and Terminal Fundamental Frequencies (F_0)

Sentence		W1	Duration (ms)				Terminal F_0 (Hz)
		<u>six</u>	[s]	[I]	[k]	Total	
1.	a.		75	46	51	172	98
	b.		135	62	92	289	62
		<u>kiss</u>	[k^h]	[I]		Total	
2.	a.		46	51		97	86
	b.		63	87		150	53
		<u>miss</u>	[m]	[I]		Total	
3.	a.		39	50		89	82
	b.		106	91		197	50
		<u>worse</u>	(not segmentable)			Total	
4.	a.					151	87
	b.					233	54

Definitions: [s] = fricative noise
[I] = voiced portion
[k] = silent closure interval
[k^h] = release burst and aspiration
[m] = nasal murmur
terminal F_0 = average F_0 of the last three complete pitch periods

The absence of any systematic prosodic effects then presumably has to do with the presence of a constant [s] noise between the prosody-carrying portion of W1 and the critical silent interval. This constant signal portion may have acted as a buffer against prosodic influences, and if so, it must be concluded that these influences are quite local in nature. In earlier studies using the fricative-affricate contrast, the distinctive prosodic information continued right up to the beginning of the silence. As was already pointed out above, there was little doubt that the [s] noise, had it been allowed to vary according to its natural production characteristics in clause-final and non-clause final position, would have had a strong influence on subjects' likelihood of reporting labial stop percepts. Such an effect would have been expected on the basis of fricative noise duration alone (Repp, 1984b; Summerfield, Bailey, Seton, & Dorman, 1981).[6]

4. Summary and Conclusions

In the present study it was attempted to create a perceptual discontinuity at the point of a critical silent interval by purely linguistic means in a relatively natural speech processing situation. The effect of word boundaries was studied, as well as the effects of (slightly removed) prosodic and syntactic breaks, following earlier studies by Dechovitz (1979, 1980, 1981), Rakerd et al. (1982), and Price and Levitt (1983).

There was a clear effect of introducing a word boundary. Although this effect was confounded with the presence vs. absence of preceding word context and therefore must be interpreted with care, it does suggest the possibility that within-word silence is more tightly integrated into the speech stream than is between-word silence. The reason for this may lie in subjects' expectations based on experience with real speech, in which interword intervals tend to be less reliable indicators of phonetic distinctions than intraword silences.

In contrast to several previous studies, there were no effects of prosodic discontinuity. The most likely explanation for this is the fact that the fricative noise immediately preceding the critical silence was not allowed to vary, so that the distinctive prosodic information ended 78 ms before the silent interval. If this interpretation is correct, it indicates that prosodic effects of the kind demonstrated by Price and Levitt (1983) and Rakerd et al. (1982) are extremely local in character and are probably caused by the duration of the acoustic segment preceding the silence, which acts as a secondary stop manner cue. Similarly restricted effects have been observed in related experiments on the perception of vowel duration in sentence context (Luce & Charles-Luce, 1985; Nooteboom & Doodeman, 1980) and on the perceptual consequences of varying speaking rate (e.g., Summerfield, 1981). Rather than constituting a direct influence of suprasegmental variation on segmental perception, these effects may be mediated by changes in local acoustic signal properties serving as segmental cues.

There were no consistent effects of syntactic structure _per se_ on phonetic perception. The anomalous results for one sentence pair were probably due to a semantic bias. These negative results confirm the conclusions of Price and Levitt (1983) and cast further doubt on the replicability of Dechovitz's (1979, 1980, 1981) unpublished findings showing a "purely syntactic" effect on phonetic perception. It seems likely that syntactic processes operate exclusively at a level beyond that of segmental phonetic classification.

References

Dechovitz, D. (1979). Effects of syntax on the perceptual integration of segmental features. In J. J. Wolf & D. H. Klatt (eds.), _Speech communication papers presented at the 97th Meeting of the Acoustical Society of America_ (pp. 319-322). New York: Acoustical Society of America.

Dechovitz, D. R. (1980). Constraints imposed on phonetic processing by syntax. _Journal of the Acoustical Society of America_, _67_, Suppl. 1, S99.

Dechovitz, D. R. (1981). Phonetic integration as a potential test for controversial surface structures. _Journal of the Acoustical Society of America_, _69_, Suppl. 1, S115.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. _Journal of the Acoustical Society of America_, _65_, 1518-1532.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. _Perception & Psychophysics_, _27_, 343-350.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. _Journal of Experimental Psychology: Human Perception and Performance_, _10_, 526-540.

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. _Journal of Experimental Psychology: Human Perception and_

Performance, 6, 110-125.
Luce, P. A., & Charles-Luce, J. (1985). Perception of word-final voicing in two sentence contexts. Paper presented at the 109th meeting of the Acoustical Society of America, Austin, Texas, April 1985.
Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. Cognitive Psychology, 10, 29-63.
Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. Cognition, 8, 1-71.
Massaro, D. W., and Cohen, M. M. (1983). Phonological context in speech perception. Perception & Psychophysics, 34, 338-348.
Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. Journal of the Acoustical Society of America, 67, 276-287.
Pastore, R. E., Szczesiul, R., & Rosenblum, L. (1984). Does silence simply separate speech components? Journal of the Acoustical Society of America, 75, 1904-1907.
Price, P. J., & Levitt, A. G. (1983). The relative roles of syntax and prosody in the perception of the /š/-/č/ distinction. Language and Speech, 26, 291-304.
Rakerd, B., Dechovitz, D. R., & Verbrugge, R. R. (1982). An effect of sentence finality on the phonetic significance of silence. Language and Speech, 25, 267-282.
Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.
Repp, B. H. (1984a). The role of release bursts in the perception of [s]-stop clusters. Journal of the Acoustical Society of America, 75, 1219-1230.
Repp, B. H. (1984b). Effects of temporal stimulus properties on perception of the [sl]-[spl] distinction. Phonetica, 41, 117-124.
Repp, B. H. (1985). Perceptual coherence of speech: Stability of silence-cued stop consonants. Journal of Experimental Psychology: Human Perception and Performance, 11, 799-813.
Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 4, 621-637.
Summerfield, Q. (1981). On articulatory rate and perceptual constancy in phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 7, 1074-1095.
Summerfield, Q., Bailey, P. J., Seton, J., & Dorman, M. F. (1981). Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split.' Phonetica, 38, 181-192.

Footnotes

[1]However, the effects studied here do not require phonetic ambiguity, as do most other contextual effects in speech perception. Rather, these purely structural effects, if extant, should disrupt the perceptual contribution of closure silence even at its optimal, least ambiguous setting. (That there is often some ambiguity even at that setting is due to the fact that closure duration is only a secondary cue to stop manner; see Repp, 1984a.)

[2]In the author's judgment, word-final [s] noises were not acceptable as word-initial segments, whereas the word-initial [s] seemed acceptable both as a word-initial or word-final segment. In any case, in the sentences and word pairs lexical and semantic constraints were assumed to exert sufficient pressure on listeners to consider the [s] as W1-final, even if its acoustic characteristics were more appropriate for a word-initial position.

[3]Except for a small reversed effect for sentence 2, $F(1,9) = 10.38$, $p < .02$, which interacted strongly with one of the two W2 factors, $F(1,9) = 69.05$, $p < .0001$, being due entirely to the clause-initial version of W2. The reason for this interaction is not known.

[4]It is conceivable that potential effects of syntactic structure were attenuated in the sentence test because the repetition of the same sentences and listeners' knowledge of the critical phonetic contrast gave rise to selective listening strategies. However, the original positive findings of Dechovitz (1979, 1980, 1981) were obtained with even more repetitive materials, and at least some degree of attention to preceding context was maintained by the requirement of key word responses in the sentence test. Moreover, in the pretest both sentences 1 and 4 showed an effect of syntactic structure at the longer closure duration ("b" responses were given only when there was no syntactic break preceding W2), whereas sentences 2 and 3 showed no effects. Thus there was no syntactic effect in the semantically unbiased sentences even on first hearing.

[5]Price and Levitt (1983) found no prosodic effect in a cross-splicing experiment similar to the present one, but this may have been due to an unusually clear-cut phonetic contrast cued by a small amount of closure silence, a situation that was not duplicated here.

[6]Two additional stimulus variables were lodged in the critical W2 word: one contrasting the strong and weak versions of W2, and the other contrasting the clause-initial and non-clause-initial versions. The effects of these factors followed a highly varied and token-dependent pattern of results and are of only marginal interest here. Details may be obtained from the author.

PERCEPTION OF THE [m]-[n] DISTINCTION IN CV SYLLABLES*

Bruno H. Repp

Abstract. The contribution of the nasal murmur and the vocalic formant transitions to perception of the [m]-[n] distinction in utterance-initial position preceding [i, ɑ, u] was investigated, extending the recent work of Kurowski and Blumstein (1984). A variety of waveform-editing procedures were applied to syllables produced by six different talkers. Listeners' judgments of the edited stimuli confirmed that the nasal murmur makes a significant contribution to place of articulation perception. Murmur and transition information appeared to be integrated at a genuinely perceptual, not an abstract cognitive, level. This was particularly evident in [-i] context, where only the simultaneous presence of murmur and transition components permitted accurate place of articulation identification. The perceptual information seemed to be purely relational in this case. It also seemed to be context-specific, since the spectral change from the murmur to the vowel onset did not follow an invariant pattern across front and back vowels.

In a recent study on the perceptual integration of nasal murmur and vocalic formant transition cues to place of articulation of nasal consonants, Kurowski and Blumstein (1984)--henceforth, K&B--showed that not only did both cues contribute to the perception of the [m]-[n] distinction, but also that their contributions were nearly equal. Their materials were 50 CV syllables uttered by a male speaker of American English, five tokens each of [m,n] followed by [i,e,ɑ,o,u]. Portions of these syllables were presented to listeners as follows: (1) the full murmur (up to the point of consonantal release); (2) the full vowel[1] (i.e., the stimulus portion following the release, which included initial formant transitions); (3) the last six pitch pulses of the murmur; (4) the first six pitch pulses of the vowel; and (5) the last three pulses of the murmur followed by the first three pulses of the vowel (i.e., the six pulses surrounding the release). The principal findings were that (a) the full murmur and the full vowel were about equally informative when presented separately (about 80 percent correct place of articulation identification); (b) shortening of these stimulus portions to only six pitch pulses led to a nonsignificant decrease in identification scores (about 77 percent correct); and (c) scores were highest for stimuli that included both the end of the murmur and the beginning of the vowel (89 percent correct).[2]

Although it was known from earlier studies that the vocalic formant transitions are strong cues to place of articulation in nasal consonants

---

*In press. Journal of the Acoustical Society of America.
Acknowledgment. This research was supported by NICHD grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. Some results were reported at the 108th meeting of the Acoustical Society of America in Minneapolis, MN, October 1984.

(e.g., Larkey, Wald, & Strange, 1978; Liberman, Delattre, Cooper, & Gerstman, 1954) and also that nasal murmurs in isolation can be identified at levels better than chance (Malécot, 1956; Nakata, 1959), K&B were the first to systematically compare identification of the two stimulus components in isolation and in combination. Their study contrasts with previous work by Malécot (1956), Nord (1976), and Recasens (1983), who used various combinations of conflicting murmurs and transitions to assess their relative contributions. In such stimuli, the transitions almost always emerge as the dominant place of articulation cue. K&B point out that this result could be due to artificial spectral discontinuities occurring at the splicing point, although the mechanism that would lead to perceptual dominance of the transitions over the murmur in such a situation has not been defined. (See Tartter, Kat, Samuel, & Repp, 1983, for a similar argument concerning the perception of stop consonant place of articulation in VCV stimuli.) In any case, K&B avoided this possible problem by combining only murmurs and transitions deriving from the same utterance. This, however, resulted in an ambiguity of their results that they acknowledge: The murmur and the transitions could act as independent cues that are combined at some higher level of processing (cf. Massaro & Oden, 1980; Repp, 1982), or the murmur and the transitions might be integrated at an early perceptual level and thus might constitute a single effective cue. This second possibility was favored by K&B on grounds of parsimony and because it is more compatible with the search for invariant properties that Blumstein and her associates are engaged in (e.g., Blumstein & Stevens, 1979, 1980; Lahiri, Gewirth, & Blumstein, 1984). These two hypotheses may be called the multiple-cue (or late integration) and single-cue (or early integration) hypotheses, respectively.

The present experiment addressed several issues relevant to these hypotheses, as applied to nasal consonant perception, thereby extending the work of K&B. Although the study was mainly an attempt to replicate the results of K&B using a larger variety of test utterances and conditions, some of the conditions were novel and explored the nature of the perceptual integration process and the role of dynamic stimulus information.

Although K&B's study was carefully conducted and incorporated five different vowel contexts, it had two methodological limitations. One is the use of a single talker: The surprisingly high identification scores for isolated murmurs could have reflected a peculiarity of his articulation. The other feature is that the subjects were permitted to respond with "b" and "d" (rather than "m" and "n") to the isolated vowel portions. While these stimuli indeed lacked nasal manner cues, the use of different response categories introduced a confounding factor. If it were the case that listeners applied slightly different criteria in place of articulation decisions for oral and nasal stop consonants (see Miller, 1977), then the scores for isolated vowel stimuli--containing acoustic information appropriate for nasal stops but being labeled as oral stops--may have been artificially depressed. It seemed important to rule out both of these possibilities, for they endanger the principal results and conclusions of K&B. The present study achieved this (1) by using six different talkers, at the price of sacrificing the assessment of within-talker variability and of using only three vowel contexts, and (2) by requiring a forced choice between "m" and "n" for all stimuli, at the price of creating a more restricted response situation.

In addition to these methodological changes, the present study expanded the range of techniques employed to assess the nature and distribution of the place of articulation information for nasal consonants. Five different

waveform editing techniques were used, each with a number of gradations: (a) Progressive truncation from the beginning of the syllable; (b) Progressive truncation from the end; (c) Extraction of brief segments from the vicinity of the consonantal release; (d) Replacement of corresponding segments in the intact syllable with noise; (e) Elimination of dynamic spectral variation in short excerpts.

These techniques complemented each other in mapping out the temporal distribution of the acoustic cues that enable listeners to distinguish [m] and [n] in utterance-initial position. In particular, they provided additional information about the relative importance of perceiving the spectral change from the murmur into the vowel. Although K&B did not emphasize this point, it is clear from their approach that they considered spectral change as the basis for an invariant property associated with place of articulation (cf. Lahiri et al., 1984). The gradual truncation conditions (a and b) assessed how much of the murmur or the vowel is needed to maintain accurate perception, and whether there is an abrupt drop in performance when one of these portions is removed altogether. The extraction condition (c) tested whether performance would be better for brief excerpts straddling the release (the point of maximal spectral change) than for excerpts of the same duration from within the murmur or vowel, thus partially replicating K&B. Conversely, the replacement condition (d) asked the same question by selectively replacing acoustic segments from within the syllable with noise, the prediction being that performance would be hurt most when the replaced segment included the point of release. An additional question of interest in that condition concerned subjects' ability to integrate murmur and vowel information across an intervening noise, allowing for the possibility of some form of perceptual restoration of the missing acoustic information (cf. Samuel, 1981; Warren, 1970, 1984; Whalen & Samuel, 1985). The final condition (e) explored the role of dynamic spectral change in the murmur and the vowel by concatenating steady-state murmur and vowel segments. The perceptual data were supplemented by an acoustic analysis of the stimuli, to determine any invariant correlate of the [m]-[n] contrast.

I. METHODS

A. Talkers and Recording Procedure

Six talkers, three males (AA, TG, SS) and three females (CG, SM, BT), participated, all native speakers of American English. AA is an experienced phonetician in his late fifties; the others are investigators or graduate students under 40 years of age.

The talkers were asked to produce the syllables [mɑ, mi, mu, nɑ, ni, nu] twice in that order, with similar intonation for all syllables. The recording session was deliberately informal and permitted a variety of speaking styles. The syllables were recorded using a Sennheiser microphone, placed approximately 10 inches from the talker's mouth, and a high-quality tape recorder.

B. Stimuli and Test Sequences

One good token of each syllable was selected from each talker's productions. The basic stimulus set thus consisted of 36 syllables (6 talkers x 6 utterances). These syllables were low-pass filtered at 4.9 kHz, digitized at a 10 kHz sampling rate, and stored in separate computer files. Using a

waveform editing program, seven markers ("cutpoints") were subsequently placed in each file, as illustrated in Figure 1. The marker labeled "0" was placed at the onset of the first pitch pulse following the point of release. This point was defined as a visible increase in high-frequency components in the oscillogram, as is clearly illustrated in Figure 1; it could be located without difficulty in all tokens. In some syllables, it fell within a glottal cycle, as illustrated in the lower panel of Figure 1. (This occasional contamination of what was, by definition, the last pitch pulse of the murmur must be kept in mind when interpreting the data.) Owing to the necessity of placing the markers at zero crossings, different criteria for the onset of a pitch period were used for male and female utterances, as shown in Figure 1: In male waveforms, the marker was placed at a downgoing zero crossing, but in female waveforms, where the downgoing slope was often very steep, it was placed at the preceding upgoing zero crossing. No perceptual consequences of this difference were expected.[3]

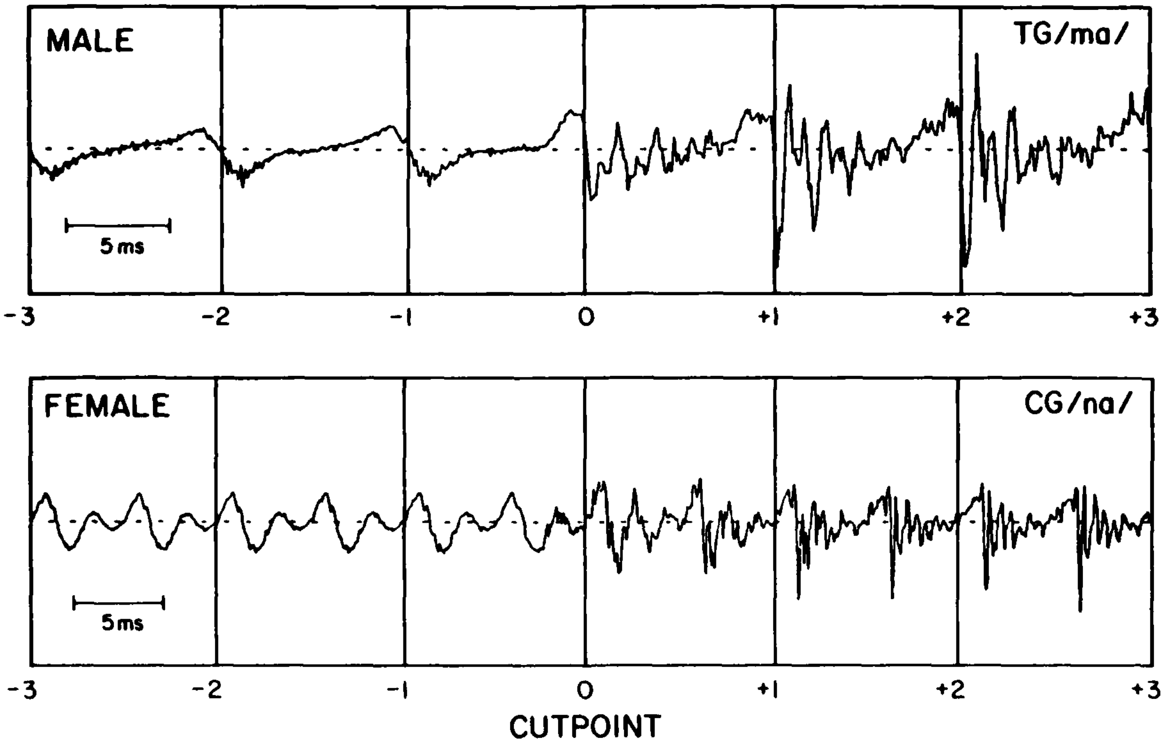


Figure 1. Central portions of the waveforms of [mα] produced by a male talker (TG) and of [nα] produced by a female talker (CG). The figure illustrates the placement of cutpoint markers.

The other six markers, labeled -3, -2, -1, +1, +2, and +3, were placed at corresponding locations at the onsets of the three preceding and following pitch periods in male utterances. In female utterances, with their higher fundamental frequencies, the pitch periods were treated in pairs, as illustrated in Figure 1. (Thus the -3 marker, for example, was placed six pitch periods before the release.) The average durations of the intermarker

intervals, calculated over the -2 to +2 range, and the corresponding fundamental frequencies for the six talkers were as follows: 10.3 ms, 97 Hz (AA); 8.9 ms, 112 Hz (TG); 10.4 ms, 97 Hz (SS); 10.4 ms, 193 Hz (CG); 10.9 ms, 183 Hz (SM); 10.5 ms, 190 Hz (BT). In the following discussion, the intermarker interval (also referred to as "segment duration") will be assigned a nominal duration of 10 ms.[4]

The set of 36 waveforms, with cutpoint markers in place, was used to generate a variety of test sequences. There were five test tapes, corresponding to the five parts of the experiment (a-e). Each tape contained between 5 and 8 test sequences. Each test sequence consisted of a single randomization of the 36 syllables, with various modifications as described below. The interstimulus interval was 3 s; there were longer pauses between test sequences.

(a) Truncation from the beginning ("Vowels"). This tape contained 8 test sequences. The first sequence contained the unaltered syllables, and the subsequent sequences presented the stimuli starting at cutpoints -3, -2, -1, 0, +1, +2, and +3, in that order.

(b) Truncation from the end ("Murmurs"). This tape also contained 8 test sequences. The first sequence contained the unaltered syllables, and the subsequent sequences presented the stimuli up to cutpoints +3, +2, +1, 0, -1, -2, and -3, in that order. It should be noted here that the murmur portions varied widely in duration, ranging from 46 ms to 223 ms, with an average duration of 103 ms.[5] Thus there was little left of some murmurs in the most extreme truncation condition.

(c) Extraction of brief segments ("Excerpts"). This tape contained 7 test sequences presenting the following excerpts: -3/+3 (i.e., from cutpoint -3 to cutpoint +3), -2/+2, -1/+1, -2/0, 0/+2, -3/-1, and +1/+3. Thus the duration of the stimuli was about 60 ms in the first sequence, 40 ms in the second sequence, and 20 ms in the remaining sequences. The segments in sequences 1-3 straddled the release, whereas those in sequences 4-7 came from within the murmur (4,6) or the vowel (5,7).

(d) Replacement of segments with signal-correlated noise ("SCN"). This tape contained 7 test sequences, with the replaced excerpts being +1/+3, -3/-1, 0/+2, -2/0, -1/+1, -2/+2, and -3/+3 (the reverse order of the Excerpts tape). Thus, the stimuli in sequences 1-5 contained 20 ms of noise, those in sequence 6 contained 40 ms, and those in sequence 7 contained 60 ms of noise. A computer program was used to generate signal-correlated noise (SCN) from specified segments within a waveform by randomly reversing the polarity of digital sampling points with a probability of .5. This results in noise that retains the amplitude envelope of the original signal but is spectrally uniform (Schroeder, 1968). An example is shown in Figure 2. The top panels compares the waveforms of the central portions of a male [mɑ] in its original form and after the -2/+2 segment was replaced with SCN (as in test sequence 6). Below, on the left, are the smoothed Fourier spectra of the -2/0 (murmur) and 0/+2 (vowel onset) segments. Note the pronounced spectral peaks and the differences between murmur and vowel spectra. On the bottom right are the spectra of the corresponding SCN segments. It is evident that the spectral difference between "murmur" and "vowel" is erased; both the murmur-derived and the vowel-derived SCN have flat spectra with random fluctuations due to the short time window. Only the difference in absolute amplitude remains, though

it is reduced due to the conversion of low-frequency into wide-band energy, especially in the murmur segment.

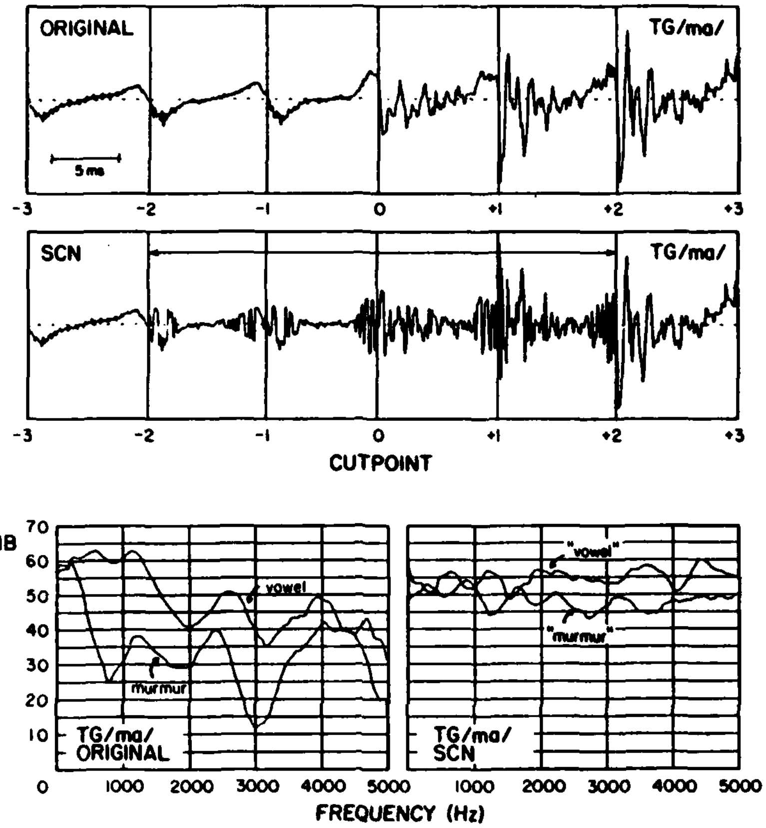


Figure 2. Central portion of the waveform of [mɑ] produced by a male talker (TG) in its original form (top panel) and after the four glottal periods between cutpoints -2 and +2 were replaced with signal-correlated noise (SCN) (center panel). The bottom panels show smoothed Fourier spectra of the "murmur" (-2/0) and "vowel" (0/+2) portions before and after replacement with SCN.

<u>(e) Elimination of dynamic spectral variation ("Static Excerpts")</u>. This final part of the experiment was exploratory in nature and included 5 test sequences. Artificial steady-state murmurs and vowels (i.e., prolonged vowel onsets) were constructed by iterating the penultimate segment (-2/-1) of the murmur and the first segment of the vowel (0/+1), respectively.[6] In the first test sequence, three repetitions of the murmur segment (i.e., three male or six female pitch pulses) were followed by three repetitions of the vowel segment. In sequences 2 (murmurs) and 3 (vowels), these 30-ms components were presented in isolation; and in sequences 4 (murmurs) and 5 (vowels), the static murmurs and vowel onsets were extended to 60 ms (i.e., 6 iterated segments). The artificial vowel segments, being prolonged onsets, had phonetic qualities different from the original [i, ɑ, u].

C. Subjects and Procedure

The subjects were twelve paid volunteers, mostly Yale undergraduates. Because of time constraints, two subjects could not listen to the last test tape (e). Ten of the subjects were native speakers of American English; the remaining two were native speakers of Russian and Chinese, respectively, but fluent in English. Their results did not differ systematically from those of the other subjects.

The tapes were played back at a comfortable intensity over TDH-39 earphones in a quiet room. Each subject listened to all tapes (with the two exceptions just noted) in a single session lasting about 100 minutes. The order of the Vowel, Murmur, and SCN conditions was counterbalanced across subjects. The Excerpts always followed these three conditions, and the Static Excerpts were last. This was done because the Excerpts conditions were considered the most difficult. There were short rest periods between test tapes.

Within each condition, the test sequences were presented in the order in which they had been recorded, as described above. This order generally proceeded from easy to difficult, so the earlier sequences provided practice for the later ones.[7]

The subjects' task was to label in writing each stimulus as beginning with "m" or "n"; or, if the stimulus did not sound like it contained a nasal consonant, to guess whether it was derived from a [m-] or [n-] syllable. In no case was identification of the vowel required. The subjects were told that there were a number of different talkers, that there was an equal number of [m-]-derived and [n-]-derived stimuli in each test sequence, and that all stimuli had been constructed from a single basic set. In the Vowels condition, the subjects were alerted to the fact that the stimuli in the later sequences might be perceived as beginning with an oral stop or with no consonant at all. (The correspondence of "b" and "m," and of "d" and "n," was explained.) In the Murmurs condition, the subjects were warned about the short duration of some stimuli in the later sequences. Preceding the presentation of each test tape, the stimulus manipulation was explained in nontechnical terms.

D. Statistical Analysis

The data of each condition (or a subset thereof) were subjected to two kinds of repeated-measures analysis of variance (ANOVA): In one ("across subjects"), correct responses were added up over the six talkers, and subjects constituted the random factor, with Consonant, Vowel, and Segment Duration and/or Location as fixed factors. In the other analysis ("across talkers"), correct responses were added up over the 12 (or 10) subjects, and talkers constituted the random factor, with Talker Sex as an additional fixed factor. Results from both analyses will be reported, since a genuine effect should generalize to both listener and talker populations. Of the two F values reported for each effect, the first is across subjects and the second is across talkers.

E. Acoustic Analysis

To track spectral peaks over time and from the murmur into the vowel, a standard LPC analysis (ILS package, distributed by Signal Technology, Inc.) was performed on all syllables, using 14 coefficients and a 20 ms analysis window moving in 10 ms steps. The ILS peak-picking routine was used to estimate formant frequencies. In addition, Fourier spectra of precisely specified time intervals were computed using another ILS program.

II. RESULTS AND DISCUSSION

A. Vowels

The overall results for the Vowels condition (truncation from the beginning) are shown as the solid function in Figure 3. It can be seen that identification of the full, unaltered syllables (F) was nearly perfect (99 percent correct). Elimination of the murmur (cut at 0) reduced performance to 85 percent correct, and truncation of the vowel onset reduced scores even more. However, performance was still significantly above chance when the first 30 ms of the vowel were excised (cut at +3); the remainders of the formant transitions thus still contained some usable place of articulation cues. Two aspects of these data deserve comment.

First, elimination of all but the last 20 ms of the murmur (cut at -2) reduced scores only slightly (to 96 percent correct); and the presence of only 10 ms of murmur (cut at -1) produced significantly better performance ($p <$ .001, sign test across subjects) than no murmur at all (cut at 0). Although the identifiability of 10-ms murmur segments in isolation was not tested and may conceivably be better than chance, their significant contribution is more plausibly attributed to an enhancement of transition perception than to any independent cue value of the murmur segment itself. This interpretation is consistent with K&B's hypothesis of a single integrated auditory property for nasal place of articulation. However, the advantage could also be attributed to the availability of sufficient nasal manner cues: In the author's informal judgment, the majority of the syllables cut at 0 sounded as if they began with oral stops (see also K&B's Table IV), whereas all syllables cut at -1 were perceived as beginning with nasal stops. Perception of the correct manner may have enhanced perception of the place of articulation cues.

Second, the score of 85 percent correct for isolated full vowels (cut at 0) is not unlike that obtained by K&B in their "long transitions" condition (80 percent correct), which confirms that the formant transitions provide strong but not entirely sufficient cues to place of articulation. The use of nasal rather than oral consonant responses in the present study did not seem to make a substantial difference.

These overall results need to be qualified in view of large differences among individual syllables, which are shown in Figure 4. It is evident that identification of nasal consonants was much poorer in [i] context than in [ɑ] and [u] contexts, as also observed by K&B. Identification of [mi] and especially [ni] suffered much more than the other syllables from truncation of the murmur, and at cutpoints beyond +1 the two syllables could not be discriminated at all. Thus the formant transitions, especially beyond the first pitch pulse of the vowel, did not provide salient place cues in [i] context. The sylable [ni], in addition, seemed to require at least 20 ms of

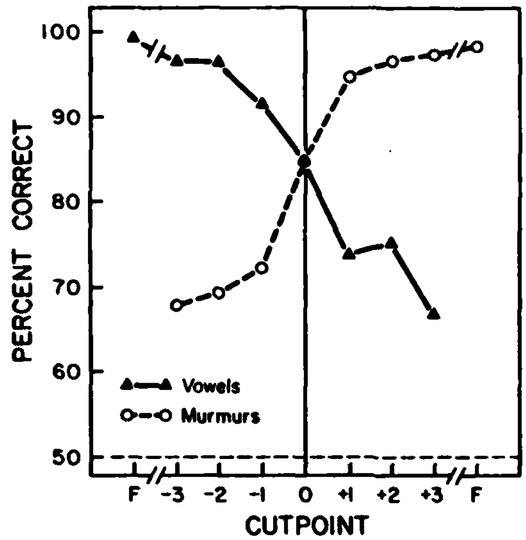


Figure 3. Percent correct identification scores as a function of stimulus duration in the Vowels and Murmurs conditions. F stands for "full syllable."

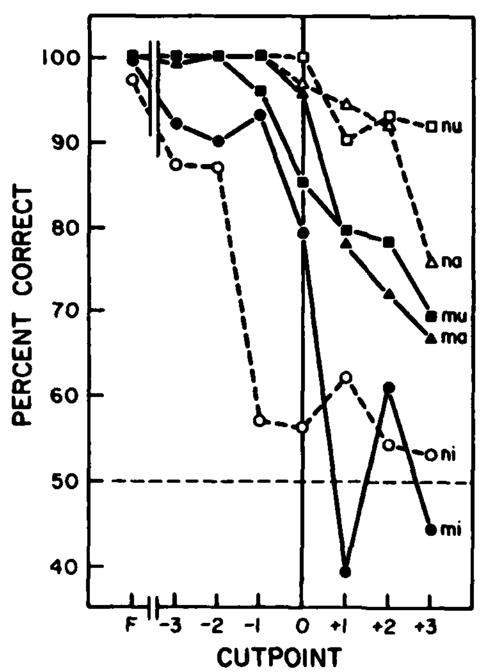


Figure 4. Individual syllable scores in the Vowels condition.

murmur to be identifiable. The data also replicate K&B's finding that [n] was identified more accurately than [m] from transitions in back vowels, while the reverse was true for the front vowel [i]. The difference in back vowel contexts can be explained in terms of transition length, reflecting distances traversed by the tongue in moving from the occlusion to the anticipated vowel configuration.

To avoid ceiling effects, only the data for cutpoints 0 and beyond (i.e., for isolated vowel stimuli) were entered into the ANOVAs, which yielded four significant effects: a main effect of Duration, $F(3,33) = 18.23$, $p < .0001$; $F(3,12) = 13.45$, $p = .0004$, reflecting the decline in performance with increasing vowel truncation; a main effect of Vowel, $F(2,22) = 67.83$, $p < .0001$; $F(2,8) = 58.79$, $p < .0001$, reflecting mainly the poorer scores for [i]; a Consonant by Duration interaction, $F(3,33) = 4.88$, $p = .0065$; $F(3,12) = 6.91$, $p = .0059$, indicating that [m] identification was hurt more by vowel truncation than was [n] identification; and a Consonant by Vowel by Duration interaction $F(6,66) = 4.41$, $p = .0008$; $F(6,24) = 2.82$, $p = .0320$, mainly due to the large advantage of [mi] over [ni] in the "0" cutpoint condition, where the Consonant by Vowel interaction described above (though it was not significant overall) was most pronounced.

Acoustic analysis of the vocalic stimulus portions revealed patterns that matched the perceptual findings. The syllables [mɑ] and [nɑ] were consistently distinguished by the second formant (F2), whose onset was 400-600 Hz higher in [nɑ] than in [mɑ]. The syllables [mu] and [nu] showed even larger differences in F2 onset, although F2 peaks could not be located reliably in three talkers' tokens of [mu]. In both [ɑ] and [u] vowels, the F2 differences persisted well beyond the first 50 ms following the release, which explains the above-chance identification of truncated vowels. The syllables [mi] and [ni], by contrast, were only minimally distinct at vowel onset. There were no indications of any difference in F2; instead, F3 and F4 onsets appeared to be somwhat higher for [ni] than for [mi]. These small differences, moreover, tended to disappear soon after the release, which explains the vulnerability of [i] vowels to truncation. All these observations are consistent with those on formant transitions in initial [b] and [d] preceding [i, ɑ, u] (Fant, 1973; Kewley-Port, 1982).

B. Murmurs

The overall results for the Murmurs condition (truncation from the end) are represented by the dashed line in Figure 3. Reading the graph from right to left, it is evident, first, that reduction of the vowel to its initial 10 ms (cut at +1) had little effect on identifiability of the consonant (94 percent correct). (Indeed, to the author these stimuli sound remarkably natural, like released nasal consonants.) This confirms that significant place-of-articulation information is located at the very onset of the vowel, immediately following the release, as has also been observed in connection with oral stop consonants (Blumstein & Stevens, 1980; Kewley-Port, Pisoni, & Studdert-Kennedy, 1983).

Complete elimination of the vowel portion (cut at 0) resulted in a clear drop in performance to 85 percent correct--the same score as for isolated vowels, and only slightly higher than K&B's score of 81 percent correct for their "long murmurs." At first blush, therefore, the results seem to replicate K&B's finding that, on the whole, isolated murmurs and vowels carry about the

same amount of place of articulation information. It must be kept in mind, however, that the last pitch pulse of the murmur was "contaminated" with incipient high-frequency energy in some syllables. Indeed, elimination of the final 10-ms segment of the murmur (cut at -1) led to a further substantial reduction in performance, to 72 percent correct. By contrast, when K&B eliminated the final pitch pulses of their isolated murmurs in a control study, performance stayed the same, which suggests that their stimuli had uncontaminated offsets. (For a possible reason, see footnote 3.) Therefore, the score of 72 percent correct is a better estimate of the intelligibility of the full isolated murmurs in the present study. Unless it is argued that the first pitch pulses of the vowel contained extra place cues due to residual nasalization and therefore should be excluded also, the conclusion must be that, overall, isolated vowels were more informative than isolated murmurs ($p < .001$, sign test across subjects). Nevertheless, identification scores for isolated murmurs were clearly above chance, which confirms K&B's general observation that these signal portions contain useful place of articulation information, probably throughout their duration.

There were large differences among individual syllables, however, which are shown in Figure 5. As in the Vowels condition, scores for [mi] and [ni]

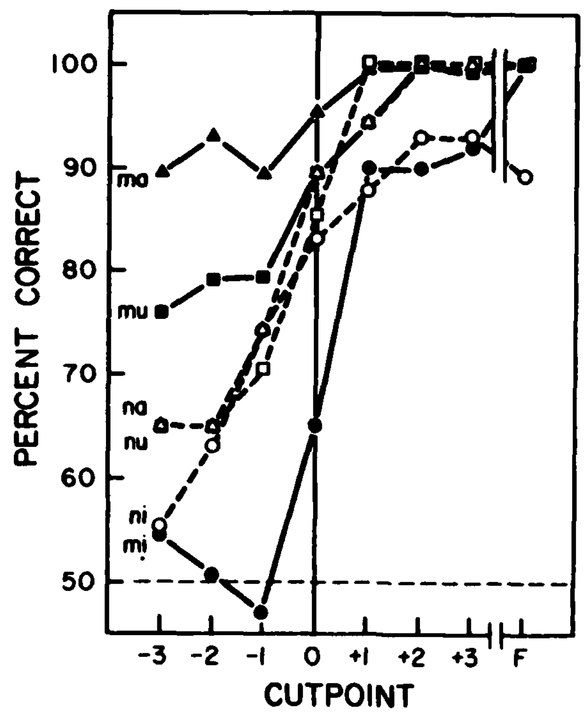


Figure 5. Individual syllable scores in the Murmurs condition.

were generally lower than those for other syllables. Thus it is not the case that the nondistinctive formant transitions in [i] are compensated for by more intelligible murmurs. Regarding the intelligibility of isolated murmurs (cut at -1, -2, -3), it seems that the differences were almost exclusively among [m] murmurs, with [m(ɑ)] best and [m(i)] worst, whereas [n] murmurs from different vocalic contexts were identified about equally well. (K&B also found that [m(i)] murmurs were much more poorly identified than [m(ɑ)] and [m(u)] murmurs, and that [n(ɑ)] and [n(u)] scores were the same; in other respects, their results were different.) Interestingly, the pattern found here is consistent with considerations from the acoustic theory of speech production: First, because of the fixation of the tongue tip during alveolar but not labial closure, lingual anticipation of the following vowel will be more evident in [m] murmurs than in [n] murmurs (see Hecker, 1962). Second, the acoustic effect of the oral shunt on the nasal murmur spectrum will be greater when the tongue body is low (as in [m(ɑ)]) than when it is high (as in [m(i)]), in proportion to the degree of coupling of the oral and nasal-pharyngeal cavities (see Kitazawa & Doshita, 1984). For these reasons, [m(ɑ)] may be expected to contain the most salient place of articulation cues, followed by [m(u)] and [n] murmurs, while the elevated tongue body during [m(i)] may in fact make this murmur more [n]-like than the [n] murmurs.

The data for uncontaminated isolated murmurs (cut at -1, -2, -3) were submitted to ANOVAs, which yielded three significant effects: a main effect of Vowel, $F(2,22) = 36.83$, $p < .0001$; $F(2,8) = 6.92$, $p = .0180$, reflecting mainly the lower scores for [-(i)] murmurs; a Consonant by Vowel interaction, $F(2,22) = 13.45$, $p = .0002$; $F(2,8) = 4.76$, $p = .0435$, reflecting the presence of a Vowel effect for [m] but not for [n] murmurs; and a Consonant by Duration interaction, $F(2,22) = 6.31$, $p = .0068$; $F(2,8) = 5.00$, $p < .0389$, which apparently derives from the fact that [n] murmurs, but not [m] murmurs, suffered from the excision of the penultimate pitch pulse (cut at -1 versus -2).[8] The lower F values in the ANOVA across talkers indicate considerable talker variability in nasal murmur spectra, a well-known phenomenon often commented on in the literature (e.g., Fant, 1960; Fujimura, 1962; Glenn & Kleiner, 1968). The unpredictable nature of that variability, as compared to the somewhat more regular scaling differences for oral resonances, may also have been responsible for the overall difference in scores between isolated murmurs and vowels in the present mixed-talker design. The subjects of K&B, of course, had to cope only with a single talker's utterances.[9]

Acoustic analysis of the nasal murmurs revealed that, in [mɑ] and [nɑ], the F2 differences observed at vowel onset were contiguous with similar F2 differences in the murmur. In other words, murmurs preceding [ɑ] generally showed distinct spectral peaks between 1 and 2 kHz, which were at least 600 Hz higher for [n] than for [m]. Although K&B did not report such a difference for their talker's [-a] murmurs, it is consistent with the acoustic theory of speech production, which predicts a lower oral resonance for [m] than for [n] (Fant, 1960; see also Saito & Itakura, 1984). Similar differences in F2 frequency tended to be present in [mu] and [nu] murmurs, though less clearly and less consistently. (See also K&B.) Differences in [mi] and [ni] murmurs were least systematic and showed large individual differences. These observations agree well with the perceptual data and the articulatory considerations presented above.

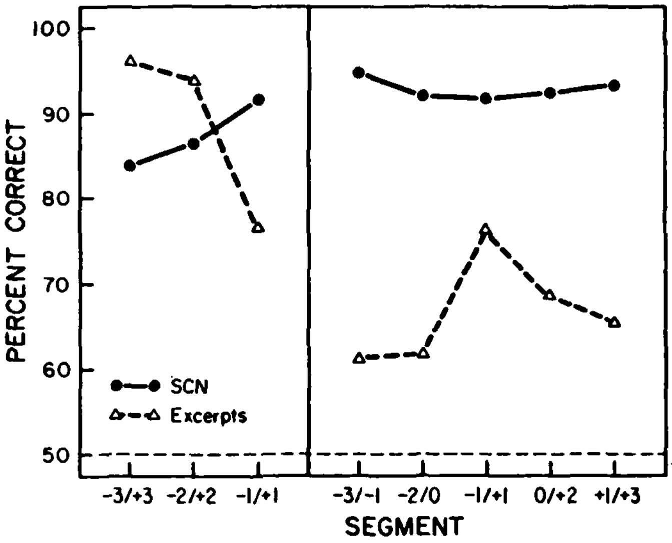


Figure 6. Percent correct identification scores in the Excerpts and SCN conditions. The left panel shows the effect of (excerpted or replaced) segment duration; the right panel shows the effect of moving a segment of constant duration across the point of release. The -1/+1 data points are duplicated in the two panels.

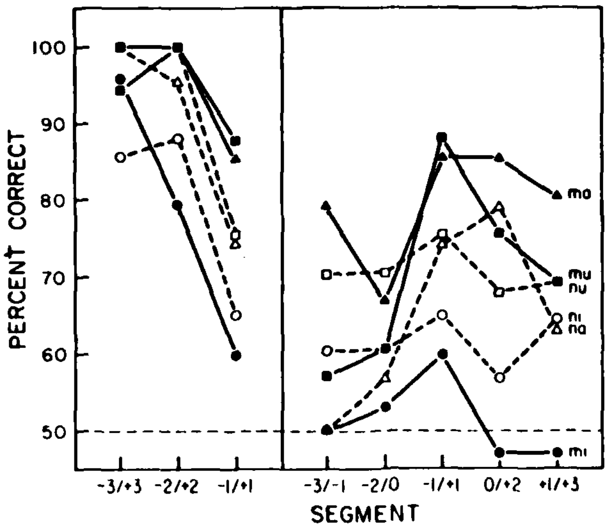


Figure 7. Individual syllable scores in the Excerpts condition.

C. Excerpts

We turn next to the Excerpts condition, which partially replicates the study of K&B. The overall results are shown as the open triangles in Figure 6. The data have been divided into two parts. On the left we see the effect of reducing the length of excerpts centered on the release from 60 to 20 ms. It can be seen that performance was quite accurate for 60- and 40-ms durations (which replicates K&B), but reduction to 20 ms resulted in a substantial decline in performance, though scores remained far better than chance. On the right in Figure 6 we see the effect of moving the location of a 20-ms excerpt across the release; the data point for -1/+1 segments is duplicated here. There was a clear peak in performance for the -1/+1 excerpts, which enclosed the release. The results thus replicate K&B's finding that identification of "mixed" excerpts is more accurate than that of equal-duration murmur or vowel ("transition") excerpts, even though the present excerpts were shorter than K&B's. Performance for 20-ms murmur excerpts (-3/-1, -2/0) was only slightly below that for vowel excerpts (0/+2, +1/+3), which is also consistent with K&B's findings.

The results for individual syllables are shown in Figure 7. Syllables including [u] and [i] all showed a tendency for 20-ms excerpt scores to peak at -1/+1; for [mɑ] and [nɑ], equivalent scores were obtained for -1/+1 and 0/+2 (vowel onset) excerpts. The rank ordering of the different syllables as vowel excerpts (0/+2, +1/+3) was not very similar to that of full isolated vowels (Figure 4: 0, +1 cutpoints), which suggests a role of the transitions beyond the initial 30 ms. The pattern for murmur excerpts (-3/-1, -2/0) was more similar to that for full isolated murmurs (Figure 5: -1, 0 cutpoints), especially for [m] murmurs.

The data for 20-ms excerpts were submitted to ANOVAs, which yielded three significant effects: a main effect of Vowel, $\underline{F}(2,22) = 20.07$, $\underline{p} < .0001$; $\underline{F}(2,8) = 25.05$, $\underline{p} = .0004$, due to the poor performance for [-i] sylables; a main effect of Location, $\underline{F}(4,44) = 4.98$, $\underline{p} = .0021$; $\underline{F}(4,16) = 5.10$, $\underline{p} = .0076$, which confirms the better performance for segments straddling the release; and a Consonant by Vowel interaction, $\underline{F}(2,22) = 21.66$, $\underline{p} < .0001$; $\underline{F}(2,8) = 6.54$, $\underline{p} = .0207$, reflecting the different Vowel effects for [m-] and [n-] syllables. The Vowel by Location interaction alluded to above (in connection with the equivalence of -1/+1 and 0/+2 scores for [-ɑ] syllables only) was marginally significant across subjects, $\underline{F}(8,88) = 2.12$, $\underline{p} = .0420$, but not across talkers.

To gain some insight into the nature of the spectral information that enabled listeners to identify place of articulation in brief excerpts straddling the release, the patterns of <u>spectral change</u> from the murmur into the vowel were examined, in the hope that they would reveal distinctive and context-insensitive patterns for [m] and [n] (cf. Lahiri et al., 1984). To quantify the change in the whole spectrum across the release, the difference between the raw Fourier spectra of the end of the murmur (-2/0) and of the onset of the vowel (0/+2) was computed for each syllable. These difference spectra are shown in Figure 8, separately for the six syllables, with the six talkers' curves superimposed. Despite considerable talker variability, fairly typical patterns of spectral change can be seen, particularly in the region between 1-3 kHz. For [mɑ] and [mu], there is less relative energy increase from the murmur into the vowel around 2-2.5 kHz than at 1 kHz, leading to a

negative slope of the difference spectrum in that region, whereas [nɑ] and [nu] difference spectra tend to have flat or rising slopes in the same region. Thus, [m] and [n] in these back vowel contexts have distinctive patterns of spectral change across the release, which largely reflect the different F2 onset frequencies and the concomitant amplitude increase in the vowel. The difference spectra for [ni], with generally rising slopes between 1 and 3 kHz, also fit this pattern; those for [mi], however, besides being highly variable, are quite different, having the most steeply rising slopes of all. The difference spectra for [mi] and [ni] differ somewhat in their slopes, which may provide a (rather unreliable) context-dependent cue for this contrast. There is no indication in these data, however, of any invariant spectral change property distinguishing [m] and [n] across all vocalic contexts.

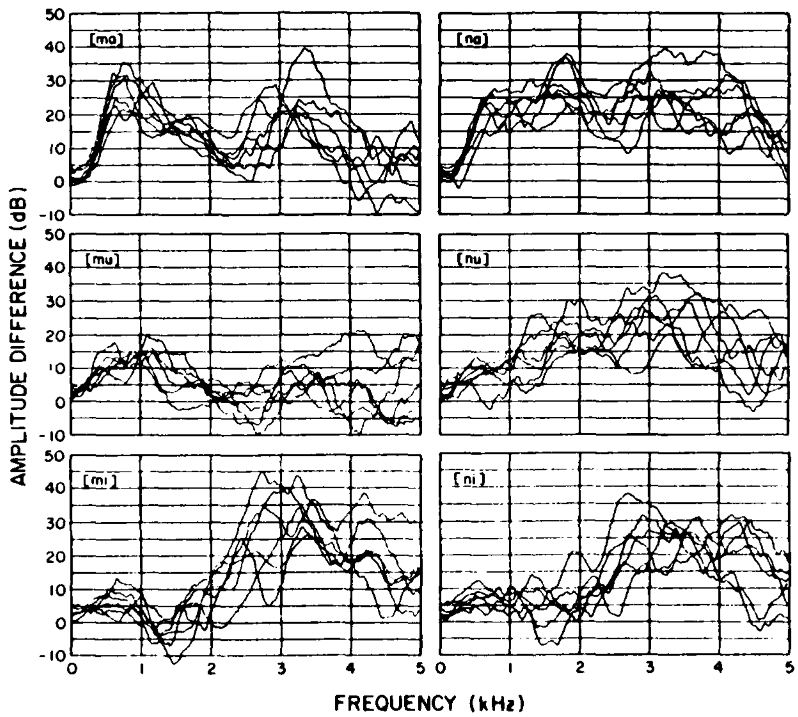


Figure 8. Raw Fourier difference spectra of 20-ms vowel (0/+2) and murmur (-2/0) segments of six syllables produced by six talkers. The difference spectra of the six talkers are superimposed in each panel.

D. Summary of Vowels, Murmurs, and Excerpts Results

The results from the three conditions discussed so far essentially confirm the findings of K&B, and they dispel any reservations about their generality across different talker populations and testing procedures. K&B's main findings--that murmurs and transitions both contribute to place of articulation identification (except perhaps in [-i] context) and that performance is best when both components are represented in a stimulus--were replicated. Their observation that murmurs and transitions in isolation are about equally identifiable was confirmed for brief excerpts, although in longer stimuli there seemed to be a certain advantage for the transitions, particularly when the vowel was [ɑ]. More significantly, perhaps, the intelligibility rank order of individual syllables was quite different for isolated murmurs and vowels, in a way that could be related to acoustic properties of the stimuli. The very poor intelligibility of both stimulus components in [-i] syllables was noted, although these syllables were identified quite well when both components were present. The spectral change across the release does not seem to provide an invariant correlate of place of articulation, though it may serve as a context-dependent cue.

E. Signal-Correlated Noise (SCN)

In this condition, it will be recalled, brief segments of the waveform in the vicinity of the release (corresponding to those presented in the Excerpts condition) were replaced with SCN, thus rendering these segments spectrally uninformative. Figure 6 shows the overall results (filled circles). Consider first the right-hand panel, where the effect of removing various 20-ms segments is shown. The question of interest here was whether replacement of the 20-ms segment straddling the release (-1/+1) would have a more detrimental effect than replacement of a 20-ms segment from within the murmur or the vowel. It can be seen that, compared to the near-perfect scores for intact syllables (Figure 3), performance was somewhat reduced in all SCN conditions, but there was no clear tendency for scores to be lowest in the -1/+1 condition. This contrasts with the clear peak obtained for the Excerpts. In the left-hand panel of the figure, which should be read from right to left for the SCN data, the effect of extending the SCN segment from 20 to 60 ms is shown. This manipulation resulted in a moderate decline in performance, but scores were still surprisingly high in the 60-ms SCN (-3/+3) condition (84 percent correct).

The scores for individual syllables are shown in Figure 9. Some striking differences are evident: [mɑ] and [nɑ] were not affected at all by SCN, not even in the most extreme condition, and [mu] and [nu] were affected only slightly in the 60-ms condition. The [mi] and [ni] syllables supplied virtually all the errors. Both of these syllables were substantially affected even by 20-ms segments of SCN, but while identification of [ni] remained above chance when the SCN segment was extended to 60 ms, identification of [mi] went to chance. There was also a difference in pattern for the two syllables: [mi], but not [ni], showed a tendency for performance to be lowest when the 20-ms SCN segment straddled the release.

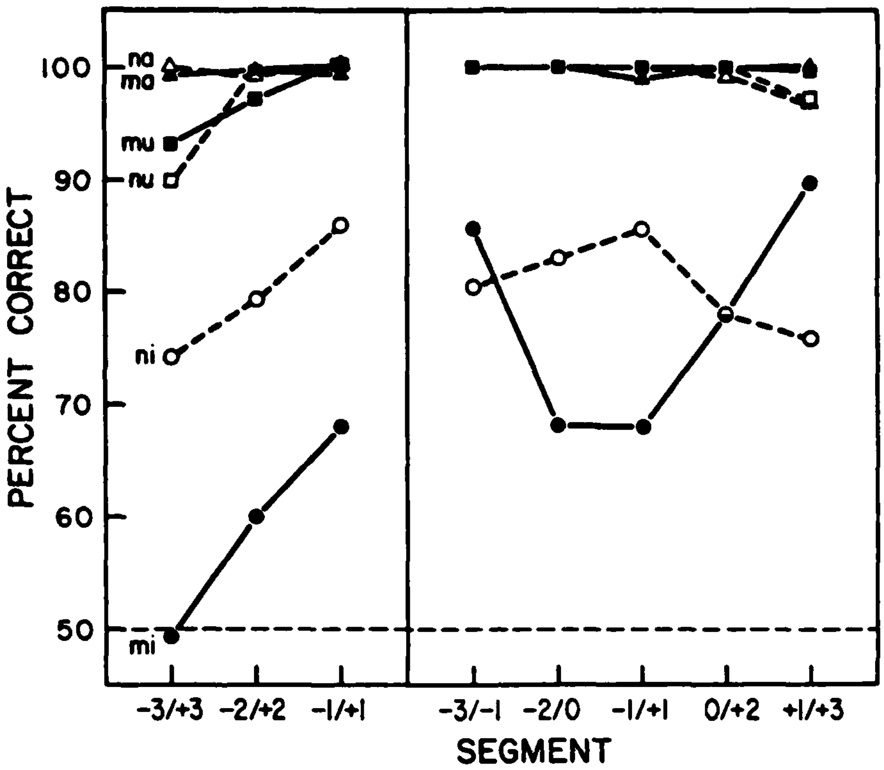


Figure 9. Individual syllable scores in the SCN condition.

Only the 20-ms data for the [mi] and [ni] syllables were submitted to ANOVAs, which yielded one significant effect: the Consonant by Location interaction just described, $\underline{F}(4,44) = 4.85$, $\underline{p} = .0025$; $\underline{F}(4,16) = 6.28$, $\underline{p} = .0031$. In the ANOVA across talkers, there was also a marginally significant effect of Talker Sex, $\underline{F}(1,4) = 8.14$, $\underline{p} = 0463$, due to higher error rates for female speech.

F. A Simple Model of "Late" Information Integration

The remarkably high performance for [-ɑ] and [-u] syllables in all SCN conditions, as well as the absence of a specific drop in performance when the 20-ms segment straddling the release was replaced with SCN (except for [mi]), raise some interesting questions about the nature of perceptual integration in these stimuli. When the murmur is immediately followed by the transitions, listeners have the opportunity to establish the single auditory property that, according to K&B's early integration hypothesis, underlies place of articulation perception. Since such auditory integration processes are likely to have a relatively short time window (a few tens of milliseconds--see

Blumstein & Stevens, 1980), they should not operate across intervening noise whose duration exceeds the integration span and which, moreover, may enter into and distort the product of integration. The excellent recognition of [mɑ] and [nɑ] when as much as 60 ms of SCN was present therefore cannot have been due to a very early integration process.

That some form of integration nevertheless took place is clear from a comparison of SCN identification scores with those for the murmur and vowel portions preceding and following the noise, obtained in the Murmurs and Vowels conditions of the experiment. For example, the average score for [nɑ] in the 60-ms (-3/+3) SCN condition was 100 percent correct, whereas that for the isolated murmur component (cut at -3) was 65 percent correct, and that for the isolated vowel component (cut at +3) was 76 percent correct. Clearly, the listeners cannot have relied on one or the other component alone; they must have combined information from the two sources in the SCN condition. (See Whalen & Samuel, 1985, for a similar result.)

It is conceivable that this integration occurred at a rather late stage in perception. Such a late integration process might evaluate each source of information separately and then combine the results according to some probabilistic rule, much as proposed by Massaro and Oden (1980). The well-known model of these authors, however, is formulated for designs in which two or more cues are varied factorially; it cannot be applied directly to experiments in which two cues are presented separately and in combination. A very simple "late integration" model may be devised for this situation, however, based on the following assumptions: (a) A stimulus component either provides "correct" information for the phonetic segment intended by the talker, with a certain probability, or it provides none at all, in which case the listener makes a random guess (i.e., we exclude the possibility that a cue reverses polarity due to some manipulation). (b) When two components are present, a listener will respond correctly when either component provides correct information (i.e., it is not necessary that both of them do). This second assumption is conservative and predicts a maximal benefit from the presence of two independent sources of information, thus counteracting the hypothesis to be tested shortly, viz., that actual performance is even better than predicted by this model.

Expressed in more formal terms, the probability of giving a correct response to an isolated murmur component is assumed to be

$$P_m = p_m + .5(1 - p_m) \;, \qquad (1a)$$

and similarly for an isolated vowel component,

$$P_v = p_v + .5(1 - p_v) \;. \qquad (1b)$$

P_m and P_v are the observed response proportions, while p_m and p_v are probabilities reflecting the information content of each component. We wish to predict from P_m and P_v the correct response proportion when both components are present, P_{mv} Since an incorrect response will result only when neither component is informative, and then only in half of the instances because of random guessing between two aternatives, we find that

$$P_{mv} = 1 - .5(1 - p_m)(1 - p_v) \;. \qquad (2)$$

From equations 1a and 1b we can derive that $p_m = 2P_m - 1$, and $p_v = 2P_v - 1$, which may be substituted into equation 2. After some simplification, this yields

$$\hat{P}_{mv} = 1 - 2(1 - P_m)(1 - P_v) , \qquad (3)$$

which is the sought-after prediction formula.

We can now attempt to predict the results for murmur-vowel stimuli from the results for isolated murmur and vowel components (even though averaging of scores over subjects and talkers may introduce some distortion in the calculations). If the obtained scores, P_{mv}, match the predicted scores, $\hat{P}_{mv}$, we may conclude that integration of murmur and vowel information took place at a late stage. If $\hat{P}_{mv}$ scores exceed P_{mv} scores, on the other hand, some more direct, more "perceptual" kind of integration would be indicated.

Table 1 presents the difference scores, $P_{mv} - \hat{P}_{mv}$, for individual syllables in four conditions: full syllables (scores averaged over the replications of this test in the Murmurs and Vowels conditions) and SCN syllabes with 20 ms, 40 ms, and 60 ms of noise centered over the release (-1/+1, -2/+2, -3/+3). The P_m and P_v scores for the predictions come from the Murmurs (0, -1, -2, -3) and Vowels (0, +1, +2, +3) conditions, respectively. A positive difference score thus means that the obtained score exceeded the predicted one. It is evident from Table 1 that the difference scores are mostly positive and quite large in some instances. (Exceptions are full [-ɑ] and [-u] syllables, for which predicted scores were very high, and [mi] in the SCN conditions, for which all scores were very low. The large difference score for [mi] in the -1/+1 condition may be an abnormality, since below-chance performance was predicted.) Moreover, there is no clear trend for difference scores to decrease as the SCN increased in duration. This leads to the tentative conclusion that some form of early perceptual integration did occur, not only when murmur and vowel followed immediately upon each other (as hypothesized by K&B), but also when as much as 60 ms of noise intervened.

Table 1

Percentage Differences Between Obtained Scores P_{mv} and Predicted Scores $\hat{P}_{mv}$ for Individual Syllables in Four Conditions.

Conditions	Syllables					
	[mi]	[ni]	[mɑ]	[nɑ]	[mu]	[nu]
Full	14	8	0	1	3	0
SCN (-1/+1)	33	6	4	3	9	5
SCN (-2/+2)	-2	13	4	5	6	4
SCN (-3/+3)	1	15	6	17	8	-4

What could account for this perceptual integration across such a reatively wide interval? One possibility is that the murmur spectrum somehow survives in auditory memory, not being masked by the following noise, so that auditory integration still occurs when the vowel begins. Another possibility

is that the acoustic information replaced by the noise is somehow reconstituted in the listener's perceptual system from long-term knowledge of acoustic-phonetic properties of speech, in a manner akin to the "phonemic restoration" phenomenon (see Samuel, 1981; Warren, 1970, 1984; Whalen & Samuel, 1985), so that perceptual integration of the filled-in information with the actual input becomes possible. Yet other possibilities, of course, are that the simple model applied in this section is based on faulty assumptions, or that isolation of stimulus components changes their aoustic properties in ways that make predictions of the sort attempted here inappropriate. We will return to this last issue in the General Discussion.

Static Excerpts

The final condition of the experiment, it will be recalled, examined the contribution of dynamic spectral change within the murmur and particularly within the vowel (the formant transitions) by presenting steady-state signal components generated by iterating one (male) or two (female) pitch periods. At the same time, the design of the Static Excerpts condition replicated rather closely the conditions employed by K&B. The questions of interest were whether concatenation of a static murmur and a static vowel onset would enable listeners to identify the nasal consonants accurately, and how scores in that condition would compare with those for stimuli containing dynamic changes and those for isolated static murmurs and vowels.

The results are presented in Table 2. Looking first at the 3M+3V results, we see that the average score for these 60-ms murmur-vowel stimuli (89 percent correct) was only slightly lower than that for the corresponding dynamic (-3/+3) stimuli in the Excerpts condition (96 percent correct). Moreover, it is immediately evident that this reduction was entirely due to the syllable [mi], which could not be identified at all in static excerpts. Identification of the other five syllables was basically unaffected by removal of dynamic information. This result indicates that the formant transitions, at least during the first 30 ms of the vowel, made no important contribution to perception of the [m]-[n] distinction. Rather, the onset spectrum of the vowel seemed to convey the distinctive information.

The poor intelligibility of [mi] in static excerpts is puzzling because the formant transition cues for that syllable seemed to be ineffective to begin with. However, the abrupt decline of [mi] scores consequent upon truncation of the first vowel segment in the Vowels condition (see Figure 4) does indicate a perceptual role of a very-short-term spectral change cue. Specifically, the vowel onset may contain a spectral transient due to the parting of the lips, whose relationship to the following vowel spectrum is perceptually important in the case of [mi]. This would also be consistent with the sensitivity of [mi] to replacement of pitch periods in the vicinity of the release with SCN, even though replacement of the -2/0 segment was even more detrimental than replacement of the 0/+2 segment (see Figure 8). Finally, the result is also consistent with the reciprocal relation of the perceptual salience of release bursts and formant transitions noted in stop consonants (Dorman, Studdert-Kennedy, & Raphael, 1977): The very ineffectiveness of the [mi] formant transitions may make even a very weak transient perceptually useful.

Turning now to the remaining four Static Excerpts tests in Table 2, it is clear that performance for these isolated steady-state murmur and vowel onset stimuli was rather poor. Scores were somewhat higher for vowel than for

murmur stimuli, and scores surprisingly declined as segment durations increased from 30 to 60 ms. This latter effect may have been due to the artificial spectral homogeneity of the stimuli, which may have become increasingly apparent to listeners as duration increased.

Table 2

Percent Correct Scores for the Static Excerpts Condition.
M = Murmur Segment (-2/-1), V = Vowel Segment (0/+1).

Conditions	Syllables						
	[mi]	[ni]	[mɑ]	[nɑ]	[mu]	[nu]	Average
3M	62	63	70	58	65	68	64
3V	38	67	73	80	80	72	68
6M	52	47	68	58	58	47	55
6V	52	55	68	67	67	60	62
3M +3V	50	92	100	98	100	95	89
$P_{mv} - \hat{P}_{mv}$	-3	16	16	15	14	13	

The data for these four tests were entered into ANOVAs with Segment Duration and Location as crossed factors, which yielded two significant effects: a main effect of Vowel, $F(2,18) = 15.20$, $p = .0001$; $F(2,8) = 8.21$, $p = .0115$, due to poorer performance for [-i] syllables; and a main effect of Duration, $F(1,9) = 6.22$, $p = .0342$; $F(1,4) = 16.66$, $p = .0151$. The main effect of Location, $F(1,9) = 3.40$, $p = .0982$; $F(1,4) = 12.79$, $p = .0232$, which compared murmur and vowel stimuli, was significant only across talkers. In the talker analysis, there was also a significant Talker Sex by Vowel interaction $F(2,8) = 4.96$, $p = .0398$: Overall, female speech accounted for more errors in [-i] and [-u] contexts and for fewer errors in [-ɑ] context than male speech.

Finally, let us compare in Table 2 the scores for isolated static components of 30 ms duration (3M, 3V) with the scores obtained when these components were concatenated (3M+3V). This comparison is analogous to that conducted by K&B, and it is clear that performance benefited enormously from the presence of both components, except in the case of [mi]. The bottom row in Table 2 shows that the increase was considerably larger than predicted by the "late integration" formula derived in the preceding section (except for [mi]), which suggests that perceptual integration, perhaps of the kind discussed by K&B, did indeed occur in these artificial stimuli.

H. Summary of SCN and Static Excerpts Results

These conditions yielded some interesting findings, which add to those of the first three conditions and of K&B. The SCN conditions and their analysis by means of a simple "late integration" model suggested that genuinely perceptual integration occurs not only when the murmur and vowel components are contiguous, but also when they are separated by as much as 60 ms of noise. While this supports K&B's general notion of a single perceptual cue, it casts

doubt on their specific hypothesis that the perceptual integration takes place at an early auditory level. The Static Excerpts results showed that, although dynamic spectral change beyond the vowel onset--such as formant movements--may contribute place-of-articulation information, this information is generally not necessary for correct identification. The syllable [mi] followed a different pattern, however, and both [mi] and [ni] were much more vulnerable to SCN than the other syllables, which suggests that the place-of-articulation information in [-i] context is of a different kind than that in [-ɑ] and [-u] contexts.

III. GENERAL DISCUSSION

The present experiment was stimulated by the recent findings of K&B that the nasal murmur and the vocalic formant transitions make about equal contributions to the perception of the [m]-[n] distinction in CV syllables. K&B used a single talker and permitted stop consonant responses when nasal manner cues were absent in the stimuli. The present study, which used six talkers and required a forced choice between "m" and "n" responses throughout, essentially confirmed the findings of K&B, although place of articulation information in the murmur seemed somewhat less salient than that in the formant transitions.

K&B hypothesized that murmur and transitions constitute a single integrated property in the auditory system, which may provide invariant perceptual information about place of articulation.[10] As to the invariant nature of this property, the present study does suggest that formant movements contribute relatively little to perception of the [m]-[n] distinction, which paves the way for an invariant measure of spectral change from the murmur to the vowel onset. Such a simple measure, however, proved to be invariant (if at all) only across the two back vowel contexts, [ɑ] and [u]; a very different criterion seems to be required to distinguish [m] and [n] in [-i] context. Indeed, it may be that spectral change cues are really important only in that context, where neither component suffices by itself.[11] It remains to be seen whether more sophisticated indices of spectral change can be found that remain more nearly invariant across different vocalic contexts.

K&B's hypothesis of a single integrated auditory property for place of articulation was supported by the present findings in so far as they suggested that the integration process does not (exclusively) take place at an abstract level of information integration. However, the isteners' apparent ability to perform such truly perceptual integration across an intervening noise (cf. Whalen & Samuel, 1985) makes it difficult to conceive of the process as a purely auditory one. At the very least, an auditory memory for spectral information must be invoked, together with an ability to reject or "listen through" noninformative noise. Although it is auditory information that is perceptually integrated, the integrative function itself should perhaps not be characterized as being auditory in nature. Indeed, it may well be specific to speech perception (Repp, 1982; see also footnote 10).

One strictly auditory process that probably does play a role in the perception of nasal consonants is short-term neural adaptation (see, e.g., Harris & Dallos, 1979). K&B (also, Blumstein & Stevens, 1979) specifically refer to Delgutte's (1980; Delgutte & Kiang, 1984) neurophysiological studies of cats, which show that a nasal murmur adapts auditory neurons in the low-frequency range, so that the response of these neurons to the onset of a

following vowel is reduced. Although there is little reason to doubt that such internal high-pass filtering of the vowel onset does occur in human listeners, it seems unlikely that this process can account fully for the perceptual integration observed. First, although short-term adaptation may extend over 100 ms or more (Delgutte, 1980; Harris & Dallos, 1979;), it may not be sufficiently strong after a 60-ms intervening noise to have much of an effect on the auditory representation of the vowel onset. Second, and more importantly, the subtraction of murmur from vowel onset spectra (Figure 8) essentially approximates (perhaps over-estimates) the high-pass filtering caused by auditory adaptation; as we have seen, no invariant property emerged from this exercise. The role of auditory adaptation nevertheless deserves continued attention: Neither K&B nor the present author took this effect into account when presenting vowel portions in isolation. One may well argue that the intelligibility of these stimulus components was reduced because not only the preceding murmur but also its auditory aftereffect had been removed. Perhaps, if the aftereffect were simulated by high-pass filtering the onsets of isolated vowels, their intelligibility would improve so much that the scores for concatenated murmur and vowel components would no longer exceed the predictions of a "late integration" model, or might even equal those for isolated vowels. This possibility is currently under investigation.

There are two reasons why high-pass filtering of vowel onsets may improve the identification of place of articulation. First, a number of studies have shown that the first formant transition may interfere somewhat with the accurate registration of higher formant transitions, so that a benefit may accrue from attenuation of F1 (e.g., Danaher & Pickett, 1975; Hannley & Dorman, 1983). Second, reduction of F1 energy may also lead to increased perception of nasal manner (e.g., Delattre, 1954), which in turn may enhance the identification of nasal consonant place of articulation. Indeed, although K&B considered place of articulation perception apart from manner perception, an important confounding factor in their study as well as in the present one was that isolated vowel stimuli were generally perceived as beginning with oral, not nasal stops. Even if the perceptual criteria pertaining to spectral correlates of place of articulation in the vowel were the same for oral and nasal stops (and they are at least very similar; see Miller, 1977), the periodic stimulus portion following a nasal stop release lacks the abrupt onset and release burst characteristics of oral stop consonants (except perhaps in [mi]). Thus, even though it may be perceived as beginning with an oral stop in isolation, it is not a "good" oral stop, and this may affect identification of place of articulation. Addition of the murmur restores perception of the correct manner class, which in itself may be responsible for at least part of the improvement in identification scores. It would be useful to dissociate manner and place perception in future research, not only by simulating low-frequency auditory adaptation but also perhaps by examining nasal consonants in the context of nasal vowels.

To conclude, while this study represents a significant extension of the work of K&B, it by no means settles all the issues raised by their work. To gain a better understanding of nasal consonant perception, future studies will have to take into account models of peripheral auditory processing, consider possible interactions of manner and place perception, and conduct a more extensive search for invariant acoustic properties.

References

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. Journal of the Acoustical Society of America, 66, 1001-1017.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. Journal of the Acoustical Society of America, 67, 648-662.

Carlson, R., Granström, B., & Pauli, S. (1972). Perceptive evaluation of segmental cues. Speech Transmission Laboratory (Stockholm), QPSR 1, 18-24.

Danaher, E. M., & Pickett, J. M. (1975). Some masking effects produced by low-frequency vowel formants in persons with sensorineural hearing loss. Journal of Speech and Hearing Research, 18, 261-271.

Delattre, P. (1954). Les attributs acoustiques de la nasalité vocalique et consonantique. Studia Linguistica, 8, 103-108.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory nerve-fibers. Journal of the Acoustical Society of America, 68, 843-857.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. Journal of the Acoustical Society of America, 75, 897-907.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. Perception & Psychophysics, 22, 109-122.

Fant, G. (1960). Acoustic theory of speech production. The Hague: Mouton.

Fant, G. (1973). Stops in CV-syllables. In Speech sounds and features (pp. 110-139). Cambridge: MIT Press.

Fujimura, O. (1962). Analysis of nasal consonants. Journal of the Acoustical Society of America, 34, 1865-1875.

Glenn, J. W., & Kleiner, N. (1968). Speaker identification based on nasal phonation. Journal of the Acoustical Society of America, 43, 368-372.

Hannley, M., & Dorman, M. F. (1983). Susceptibility to intraspeech spread of masking in listeners with sensorineural hearing loss. Journal of the Acoustical Society of America, 74, 40-51.

Harris, D. M., & Dallos, P. (1979). Forward masking of auditory nerve fiber responses. Journal of Neurophysiology, 42, 1083-1107.

Hecker, M. H. (1962). Studies of nasal consonants with an articulatory speech synthesizer. Journal of the Acoustical Society of America, 34, 179-188.

Kewley-Port, D. (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. Journal of the Acoustical Society of America, 72, 379-389.

Kewley-Port, D., Pisoni, D. B., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. Journal of the Acoustical Society of America, 73, 1779-1793.

Kitazawa, S., & Doshita, S. (1984). Nasal consonant discrimination by vowel independent features. Studia Phonologica, 18, 46-58.

Kurowski, K., & Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. Journal of the Acoustical Society of America, 76, 383-390.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants:

Evidence from a cross-language study. Journal of the Acoustical Society of America, 76, 391-404.

Larkey, L., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. Perception Psychophysics, 23, 299-312.

Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 68.

Malécot, A. (1956). Acoustic cues for nasal consonants: An experimental study involving a tape-splicing technique. Language, 32, 274-284.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. Journal of the Acoustical Society of America, 67, 996-1013.

Miller, J. L. (1977). Nonindependence of feature processing in initial consonants. Journal of Speech and Hearing Research, 20, 519-528.

Nakata, K. (1959). Synthesis and perception of nasal consonants. Journal of the Acoustical Society of America, 31, 661-666.

Nord, L. (1976). Perceptual experiments with nasals. Speech Transmission Lab. (Stockholm), QPSR 2-3, 5-8,

Recasens, D. (1983). Place cues for nasal consonants with special reference to Catalan. Journal of the Acoustical Society of America, 73, 1346-1353.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.

Saito, S., & Itakura, F. (1984). Further study on the frequency spectrum deviations between speakers. Annual Bulletin RILP, 18, 107-112.

Samuel, A. G. (1981). The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance, 7, 1124-1131.

Schroeder, M. R. (1968). Reference signal for signal quality studies. Journal of Acoustical Society of America, 44, 1735-1736.

Tartter, V. C., Kat, D., Samuel, A. G., & Repp, B. H. (1983). Perception of intervocalic stop consonants: The contributions of closure duration and formant transitions. Journal of the Acoustical Society of America, 74, 715-725.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. Science, 167, 392-393.

Warren, R. M. (1984). Perceptual restoration of obliterated sounds. Psychological Bulletin, 96, 371-383.

Whalen, D. H., & Samuel, A. G. (1985). Phonetic information is integrated across intervening nonlinguistic sounds. Perception & Psychophysics, 37, 579-587.

Footnotes

[1]K&B used the term "long transitions" for this stimulus portion. That formant transitions often extend beyond the initial 60 ms or so is illustrated by K&B's footnote 1, which reports [a] second-formant frequencies almost 300 Hz higher following [n] than [m] "around the center of the vowel well past the formant transitions" (K&B, p. 389). See also Kewley-Port (1982) for analogous observations on stop consonants.

[2]The study did not include a condition in which the full, unaltered syllables were presented for identification. By using truncated murmurs and vowels, K&B (who did not motivate this choice) presumably wanted to emphasize the concentration of place-of-articulation information around the release. However, a comparison of identification scores for full murmurs and vowels (about 80 percent correct) with those for full syllables (surely better than 90 percent correct) would have led to very similar conclusions.

[3]K&B apparently even placed their markers in the middle of glottal cycles (see their Figure 1, left-hand panel).

[4]A repeated-measures ANOVA was conducted on the intermarker intervals in the -2 to +2 range, with the factors Before/After Release, Consonant, and Vowel. There were no significant effects in this analysis, showing in particular that (1) F0 did not change abruptly at the release, and (2) F0 did not differentiate [m] and [n].

[5]A repeated-measures ANOVA was conducted on the murmur durations, with the factors Consonant and Vowel. There were no significant effects. Individual differences among talkers were considerable, however: Average murmur durations ranged from 70 to 152 ms, and standard deviations ranged from 10 to 43 ms.

[6]As pointed out earlier, the last murmur segment (-1/0) sometimes contained incipient high-frequency energy from the release; this is why the preceding murmur segment was used for iteration. The iteration of two pitch pulses in the female tokens did not result in noticeable fluctuations of timbre.

[7]This arrangement differs from that employed by K&B, who presented diverse stimuli in a single randomized sequence. The present design, with homogeneous blocks of stimuli graded according to difficulty, favored the most difficult conditions, thus working against the perceptual integration advantage resulting from the simultaneous availability of murmur and transition cues. Such an advantage was nevertheless obtained, which suggests that practice effects were negligible. Another important departure from K&B's design is the use of multiple talkers, which may have increased the difficulty of all identification tasks.

[8]An unexpected difference between male and female talkers was noted in the 0 and +1 truncation conditions, which were not included in the ANOVAs: The average scores of both conditions were 98, 98, and 94 percent correct for the three male talkers, and 90, 90, and 87 percent correct for the three female talkers. The cause of this difference is unknown. Note that there were no effects of Talker Sex for either isolated murmurs or isolated vowels.

[9]Another possibility considered was that the rather short durations of some of the murmurs employed here were responsible for the lower murmur identification scores. The average murmur duration (103 ms) was only slightly less than that in the K&B study (117 ms), but variability was much larger. However, inspection of the data revealed that, although the shortest murmurs did not receive very high scores, many long murmurs yielded scores that were equal or even poorer. Murmur duration was entered as a covariate into an analysis of covariance, which yielded results similar to the ANOVA together with a pooled regression coefficient of -0.01, indicating that murmur duration did not account for any significant variation in the data.

[10]When K&B say that "the auditory system does not treat transitions separately from the murmur" (p. 389), do they mean to imply that listeners would not be able to discriminate a stimulus with initial murmur from one in which the murmur has been deleted and the vowel onset has been modified acoustically (by some kind of high-pass filtering) to simulate the effect of the murmur on the auditory response at vowel onset? This prediction should be easy to disconfirm, for the murmur is easily detectable as a separate auditory event. If their statement is to be interpreted as meaning that, as a cue to place of articulation, the murmur and the transitions form a single integrated property, then they must mean that the integration is a speech-specific, not a general auditory function.

[11]In a perceptual study with synthetic speech, Carlson et al. (1972) found that the frequency of the second nasal formant during the murmur was critical for the [mi]-[ni] distinction. The present data offer little support for this observation.

ON THE NATURE OF MELODY-TEXT INTEGRATION IN MEMORY FOR SONGS*

Mary Louise Serafine,† Janet Davidson,†† Robert G. Crowder,†† and Bruno H. Repp

Abstract. In earlier experiments (Serafine et al., 1984) we found that the melodies of songs were better recognized when the words were those that had originally been heard with the melody than when they were different. Similarly, song texts were better recognized when sung with their original melodies. Some possible causes of this "integration effect" were investigated in the present experiments. Experiment 1 ruled out the hypothesis that integration was due to semantic connotations imposed on the melody by the words, since songs with nonsense texts yielded the same effect. Experiments 2 and 3 ruled out the possibility that the earlier results were caused by a decrement in recognition when a previously-heard component is tested in an unfamiliar context. The results support the notion of an integrated memory representation for melody and text in songs.

Songs consist of two components, melody and text, which seem to be separable in a number of ways. They can be performed, perceived, and notated separately, and in practice may be composed by different artists. At least intuitively, however, the melody and text of a song seem more tightly related than two arbitrary simultaneous events. The components of a song seem more integrated, for example, than a spoken voice with background music. These observations raise questions about the memory representation for songs and whether it consists of independent (separate) or integrated components.

In a previous study (Serafine, Crowder, & Repp, 1984) we found evidence for what we termed the integration effect--the tendency for a melody to be better recognized when the text was the one with which the melody was originally heard than when the text was different. Similarly, there was a tendency for the text to be better recognized when sung with the original melody than with a different melody. The effect for melody recognition was very robust. It held across performances by different singers and could not be eliminated voluntarily by our subjects when we instructed them to focus on melody only. We concluded that melody and text form an integrated memory representation.

Integrated memory for melody and text may explain some of the experiences that people commonly have in recalling and recognizing song components. For example, if asked to recite the words to their national anthem, many people would have to sing the song or at least rehearse it subvocally in order to

---

*Journal of Memory and Language, 1986, 25, 123-135.
†Vassar College
††Yale University
Acknowledgment. This research was supported by NSF Grant 82-19661 to R. Crowder and by NICHD Grant HD-01994 to Haskins Laboratories.

generate the words. Also, many people do not recognize even a very familiar melody if it is sung with different words. Examples are the folksong "Baa, Baa Black Sheep," which has the same melody as "Twinkle, Twinkle Little Star," and the folksong "Merrily We Roll Along," which has the same melody as "Mary Had a Little Lamb." The integration effect may also underlie the informal observation (Gottlieb, 1984) that young children are frequently unable to sing only the melody of a song if asked to replace the words with a repeated syllable such as "la." Their tendency to respond by speaking the syllable, by singing some spontaneous, unrecognizable melody, or by refusing to respond altogether may be evidence that they are unable to access the melody without its text.

Our previous study of melody-text integration employed the following method, which was similar to that used in the present experiments. Subjects heard a serial presentation of excerpts from 24 largely unfamiliar folksongs. The presentation was immediately followed by a recognition test in which two types of items were heard: (1) excerpts that had been heard in the presentation ("old songs") and (2) excerpts that had not been heard in the presentation ("new songs"). Further, new songs were of four types: (a) new melody with new words; (b) old melody with new words; (c) new melody with old words; and (d) old melody with old words that had been sung to a different melody in the original presentation ("mismatch songs"). The critical finding was that recognition of a melody (or text) under the old song condition was superior to recognition under the mismatch condition. That is, recognition of a component was better when it was paired with its original component than with a different, even if equally familiar, component. The experiments reported here were intended to evaluate two interpretations of the obtained integration effect:

<u>The semantic hypothesis</u>. The integration effect could be caused by the semantic connotation that words impose on a melody. In the more usual cases a melody may be imbued with qualities implied by the text's meaning, even if the melody on its own would not normally convey that meaning. For example, words may make some aspect of the melody particularly salient. In the present folksongs, reference to a cobbler may make a repetitive melodic pattern seem to suggest hammering; reference to a bluebird may make higher-pitched or ascending tones seem to imply flying, birdsong, etc. In some, admittedly more rare, cases the melody may overtly mimic the meaning of the words, as when a repeated eighth-note figure appears on the words "tapping at the window." More generally the text of a sea chantey, hymn, lullaby, or other stylized song could trigger (even unconscious) recognition of the special tonal and rhythmic conventions that are characteristic of such songs.

Once the melody of a song is taken to be especially related to a particular meaning, its recognition may be inhibited in the context of a different, especially if incongruous, meaning. What has suggested hammering or birdsong is less recognizable in the context of Cape Cod or an old sow's hide. The semantic hypothesis, then, accepts the reality of melody-text integration and attributes it to the semantic level. (Note that this hypothesis could account only for the integration effect in melody recognition, not for that in text recognition.)

<u>The decrement hypothesis</u>. By contrast, a second interpretation denies that the observed integration effect implies an integrated memory representation. Rather, the integration effect could be an artifact of the deleterious, distracting influence that a "wrong" component has on an already

familiar component. For example, the memory representation of a melody may be quite independent of its text, and under normal circumstances may be just as easily recognized in one condition as another. However, the mismatch condition, precisely because it contains different words, may distract or confuse subjects and depress melody recognition. In such a case the integration effect would be only an experimental artifact. The decrement hypothesis can be tested by comparing recognition of components in old songs and mismatch songs to the recognition of melodies and texts presented alone (hummed or spoken, respectively).

Experiment 1 addressed the semantic hypothesis for melody recognition. Experiments 2 and 3 addressed the decrement hypothesis for melody and text recognition, respectively. All three experiments employed the same general procedure: Subjects heard a serial presentation of folksong excerpts, followed immediately by a recognition test for melodies or words in which the items represented different combinations of old and new components. Because all three experiments employed variations of the same musical materials, these are described in some detail before the experiments proper.

General Method

Songs that we believed would be unfamiliar to the average listener were drawn from a collection of indigenous American folksongs compiled by Erdei (1974).[1] Twenty pairs of song excerpts with interchangeable melodies and texts were chosen, each excerpt consisting of the opening two to four measures of a song. (See list in appendix.) Interchangeability of words and melodies within a pair was crucial to the construction of plausible recognition foils. Thus, with two exceptions each text within a pair contained the same number of syllables, and each text contained a suitable stress pattern that would fit with either melody. The exceptions were Song Pairs 11 and 17, where one text was shorter by a syllable, and thus one syllable was sung across two tones ("slurred"), as is normally the case in the different verses of a song. (The opening "O-oh" of our national anthem is an example.)

Each pair of excerpts yielded four different songs, a total of 80. Figure 1 shows a sample pair of interchangeable melodies, and Figure 2 shows examples of the five types of test items that can be generated from each pair. These materials allowed for counterbalancing so that every presentation item could be tested against every possible test item type. Thus, natural variations among the folksongs were controlled.

In some cases minor alterations were made to the melody or text to ensure a rhythmic fit with its companion. (See appendix.) For example, "across" from one original text was changed to "cross" in our experiments (Figure 2, test item a). However, in all cases the texts and melodies were identical across presentation and test versions of a song.

The excerpts were recorded on tape, sung by a female in the alto range, at a tempo represented by one beat per second. A silent metronome was employed to ensure an accurate beat, but because of normal metric variations in the songs (e.g., "double time") the subjective tempo of the excerpts was not necessarily uniform. All songs were notated with G as the tonic, although they varied in key, mode, and starting tone. The excerpts were sung as notated, except transposed down a fifth or twelfth to the appropriate range. A pitch pipe was used to ensure starting pitch accuracy. The experimental

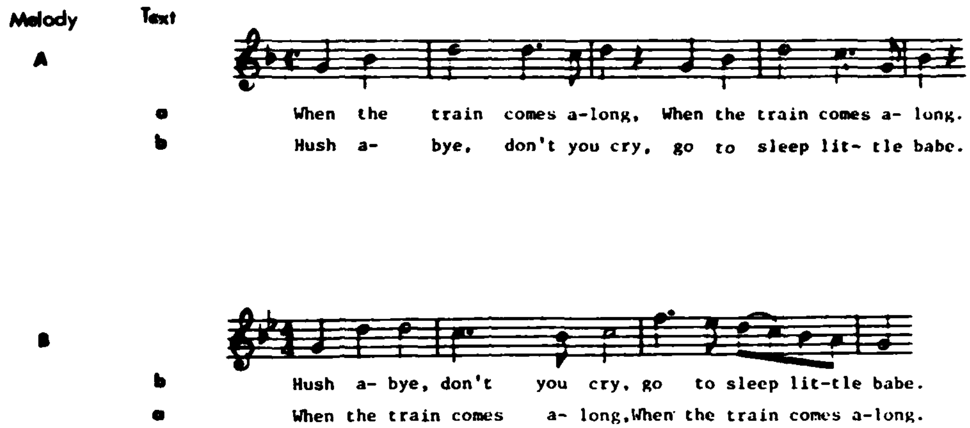


Figure 1. Sample pair of songs with interchangeable texts. (Aa and Bb denote original songs; Ab and Ba denote derivatives).

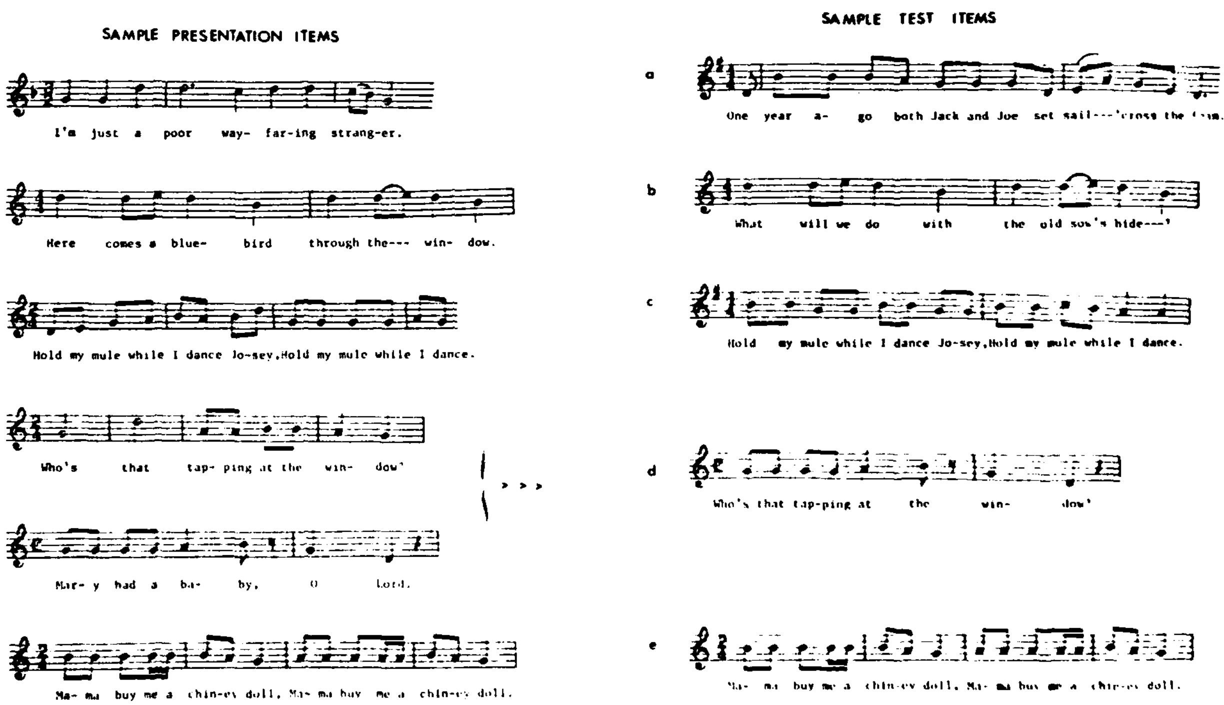



Figure 2. Sample presentation and test items. (a: new melody, new words; b: old melody, new words; c: new melody, old words; d: old melody, old words--mismatched; e: old song.)

tapes were dubbed from a master tape, with a 5-s interval of silence between presentation items and a 10-s response interval after each test item.

Experiment 1

The semantic hypothesis holds that the integration effect is due to semantic connotations that the words of a song impose on its melody. If this hypothesis were correct, the integration effect should disappear when the semantic meaning of the words is eliminated. In the present experiment subjects heard a presentation of 24 folksong excerpts in which the words had been translated into nonsense. The presentation was followed immediately by an 18-item recognition test comprising six each of the following types of items:

(a) old songs (old melody, old nonsense words) exactly as heard in the presentation;
(b) new songs (new melody, new nonsense words) that had not been heard in the presentation; and
(c) mismatch songs (old melody with old nonsense words that had been sung to a _different_ melody in the presentation).

The main prediction was that, if the semantic hypothesis were correct, melody recognition should not be better in the old song condition than it is in the mismatch condition. On the other hand, if the integration effect is due to factors other than the semantic connotation of words, then the effect should still hold when nonsense words are employed.

Method

Materials

Eighteen of the 20 pairs of interchangeable folksong excerpts listed in the appendix were used to generate presentation and test stimuli (song pairs 4 and 10 were omitted, since these each contained a song that was more frequently identified as familiar by subjects in our earlier studies). Each of the 36 texts was translated into a nonsense text by applying the following rules:

1. Vowels remain the same.
2. Consonants are interchanged according to the following list, where, if the right-listed consonant appears, it is changed into the left-listed consonant and vice versa. Phonetic classes are preserved.

B	G
K (QU, C)	T
L	Y (or F)
M	N
P	D
S (C)	F
H	J
R	W
Z	V
Sh, Th	Ch

3. Whenever necessary, license was taken with the above rule to ensure pronounceability and to eliminate accidental semantic meaning.

The following are examples of translated texts:

Original: Cobbler, cobbler, make my shoe.
Nonsense: Tog-glue, tog-glue, nate nie choo.

Original: Cape Cod girls they have no combs.
Nonsense: Tade top berf shey jaze mo tong.

The excerpts were sung and recorded on tape as described under General Method.

Design

Three parallel sets of presentation and test sequences were constructed from the set of 18 pairs of excerpts. Each set was administered to a different group of subjects. In the presentation sequences (24 items), half the excerpts were melodies with nonsense words derived from their original texts (type Aa or Bb in Figure 1), and half were melodies with nonsense words derived from the companion, interchangeable text of the pair (type Ab or Ba in Figure 1). In the test sequences (18 items), each of the three types of test items (old, new, and mismatch song) occurred six times. Further, across the three subject groups, each presentation excerpt was tested against each of the three test item types. For Test Tape 1, the three item types were assigned at random to the 18 items available (for example, old, new, and new for the first three items). Thereafter Test Tapes 2 and 3 were derived accordingly (for example, mismatch, old, old, and new, mismatch, mismatch, respectively).

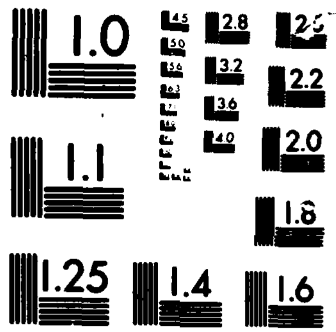
The presentation and test excerpts were generated successively from Song Pairs 1 through 20 (omitting 4 and 10), in the order listed in the appendix. Thus, the interval between each presentation item and its corresponding test item was roughly constant. Note that each of the "mismatch" test items required two presentation excerpts, since the old words of one excerpt would be paired with the old melody of another excerpt. When two such presentation excerpts were required, they immediately followed each other on the tape. (If anything this convention would inflate performance in the mismatch condition, working against the hypothesis of an integration effect.) The resulting total of 24 presentation excerpts represents the 12 excerpts necessary for the old and new test items (6 each), plus the 12 excerpts necessary for 6 mismatch items requiring two excerpts.

Procedure

Testing was conducted individually in a quiet laboratory in which presentation and test tapes were heard over loudspeakers. Subjects were instructed to listen carefully to a presentation of 24 songs that sound like folksongs, except that the words have been changed to nonsense. They were told that their "memory for the songs would be tested later," but they were given no further information. The test sequence followed immediately. For each item, subjects were asked to indicate on the answer sheet whether they had "heard that exact melody before--that is, just the musical portion" (yes or no), and to indicate the degree of confidence they felt in their judgment by marking a three-point confidence rating scale (1 = not very confident, 3 = very confident). No advance information was given about what types of items would occur on the test.

AD-A160 019 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT A M LIBERMAN JAN 86 SR-84(1985) N00014-83-K-0083 2/3

UNCLASSIFIED F/G 5/7 NL


1.0
2.8
3.2
3.6
4.0
2.2
2.0
1.1
1.8
1.25
1.4
1.6

Subjects

Thirty-seven Yale undergraduates with undetermined levels of musical training were paid to participate. The three subject groups contained 13, 12, and 12 subjects respectiveley.

Results and Discussion

Yes/no responses with confidence ratings were translated into a single rating that ranged from 1 to 6, where 1 represents very confident no (did not hear melody), and 6 represents very confident yes (did hear melody). Mean ratings for the old, new, and mismatch conditions were 4.47, 2.60, and 3.76, respectively. The results of two analyses of variance for the three conditions were significant: With subjects as the sampling variable, $F(2,72) = 51.94$, $p < .001$, and with the 18 song pairs as the sampling variable, $F(2,34) = 38.35$, $p < .001$. Post hoc analyses (Scheffé procedure) revealed that melody recognition under the old song condition (mean = 4.47) was significantly better than it was under the mismatch condition (mean = 3.76), both across subjects, $p < .01$, and across song pairs, $p < .05$.

Thus, the integration effect was confirmed with the new materials used here. Melodies were recognized better when they were paired with their original text than when paired with another, even if equally familiar text. Since this effect held when nonsense texts were used, the semantic hypothesis must be ruled out as an explanation for the integration effect. This does not imply, however, that semantic integration of melody and text never occurs. Indeed, especially in those cases where the melody directly symbolizes textual meaning (e.g., repeated eighth notes on "tapping"), integration on the semantic level seems likely. What experiment 1 does show, however, is that integration does not depend on semantic factors.

Experiment 2

Thus far, we have attributed the performance advantage in old songs over mismatch songs to a recognition superiority in the former condition. The decrement hypothesis, on the other hand, holds that the seeming advantage in old songs is due to the deleterious, distracting effect that "wrong" words have on melody recognition under the mismatch condition. If this hypothesis were correct, it could account for the performance advantage in old songs without recourse to an integrated memory representation. Perhaps the melody by itself could be recognized well without the original words, but adding new or mismatched words somehow disguises the retained melodic information.

In the present experiment, subjects heard a presentation of 24 consecutive folksong excerpts, followed by a 20-item recognition test. (Normal texts, not nonsense were used throughout.) The test items were of five types:

(a) old songs (exactly as heard in the presentation):
(b) mismatch songs (old melody with old words from a different song in the presentation);
(c) old words with new melody;
(d) hummed version of an old melody from the presentation ("old hum"); and
(e) a hummed version of a new melody that had not been heard in the presentation ("new hum").

The decrement hypothesis predicts that melody recognition when old words are present (as measured by responses to mismatch songs and old words with new melody) will be poorer than melody recognition when no words are present (as measured by responses to old hum and new hum). In essence, the hummed conditions provide a baseline against which to measure two influences. First, if there is a decrement caused by "wrong" words, then discrimination of old and new melodies should be better when they are hummed than when they are presented with old (but mismatched) words. Second, if melody-text integration has a positive or facilitative effect on melody recognition, then old (intact) songs should have a recognition advantage over old hummed melodies.

Method

Materials

The materials consisted of the same set of 20 pairs of folksongs with interchangeable texts (not nonsense) that were described previously, except that additional recordings were made by the same female alto of hummed versions of the melodies. In this experiment, two recordings done on separate occasions were made of each stimulus. This allowed for different performances to be used across presentation and test items, thus eliminating the possibility that the physical identity of old song and old hum test items (including even accidental sounds) could contribute to superior melody recognition on those items.

Design

Five parallel sets of presentation and test sequences were constructed using (in the order listed) the 20 pairs of folksong excerpts in the appendix. Each set was administered to a different group of subjects. In the presentation sequences (24 items), half the excerpts were melodies with their original texts (type *Aa* or *Bb* in Figure 1) and half were melodies with texts borrowed from their companion song (type *Ab* or *Ba* in Figure 1). In the test sequences (20 items), each of the five types of items (old song, mismatch, old words with new melody, old hum, new hum) occurred four times. Across the five subject groups each presentation item was tested against each of the five possible test item types, which were assigned by following a Latin square design. Each of the mismatch test items required two presentations, which immediately followed one another on the tape.

Procedure

The procedure was the same as that used in Experiment 1. Subjects were told to listen carefully to a presentation of 24 excerpts from simple folksongs and that their "memory would be tested later." They were not told that only melody recognition would be tested. Prior to the test they were told that items on the test would be either hummed melodies or melodies with words, but in all cases they were to disregard the words and indicate whether they had "heard this exact melody before--that is, just the musical portion." Subjects indicated yes or no on the answer sheet and gave a confidence rating.

Subjects

Forty Yale undergraduates with undetermined levels of musical training were paid to participate in the study. They were divided equally among the five presentation/test sequences.

Results and Discussion

As in the first experiment, subjects' responses were translated into ratings ranging from 1 to 6 where 1 represents very confident no (did not hear melody) and 6 represents very confident yes (did hear melody). Means for the five conditions--old songs, mismatch songs, old words with new melody, old hum, and new hum--were 4.71, 3.73, 3.21, 3.99, and 3.11 respectively. The results of analyses of variance on these means were significant both across subjects, $F(4,156) = 26.76$, $p < .001$, and across song pairs, $F(4,76) = 17.37$, $p < .001$.

Confirmation of the integration effect. Post hoc analyses (Scheffé procedure) revealed that melody recognition under the old song condition (mean = 4.71) was superior to that in the mismatch condition (mean = 3.73), both across subjects, $p < .01$, and across song pairs, $p < .01$. This confirms the integration effect found in the previous experiment.

Disconfirmation of the decrement hypothesis. For this analysis subjects' melody recognition performance was measured by difference scores with a theoretical range of -5 to +5, where incorrect recognitions were subtracted from correct recognitions (hits minus false alarms). The mean difference score when old words were present (rating for mismatch minus rating for old words/new melody) was .52. The mean difference score when no words were present (rating for old hum minus rating for new hum) was .88. The difference between these means narrowly missed the conventional level of significance, $t(39) = 1.89$, $p < .07$ (with subjects as the sampling variable), indicating that melody recognition was not significantly lower when old words were present than when no words were present. This result fails to support the decrement hypothesis, which holds that poorer recognition in the mismatch than in the old song condition (the integration effect) could be due to the fact that wrong words depress melody recognition performance. On the other hand, because the difference was close to statistical significance, we should leave this hypothesis tentatively open, the more so because melody recognition in both conditions was near chance.

The alternative hypothesis, however, that original old songs have a positive, facilitative effect on melody recognition was supported by the following results. The mean difference score when original old words were present (rating for old song minus rating for old words/new melody) was 1.49, which is significantly greater than the mean difference score when no words were present (.88 as above), $t(39) = -2.61$, $p < .02$ (with subjects as the sampling variable). Thus melodies were better recognized in the presence of their original old words than on their own, without words.

Criterion effects. To assess criterion effects, we analyzed the tendency to respond "yes, I heard the melody," whether correct or incorrect, when old words were present and in the hummed conditions. The overall rating when old words were present (mean of mismatch and old words/new melody) was 3.47, which is not significantly lower than the overall rating of 3.55 in the hummed conditions (mean of old hum and new hum). The Scheffé procedure yielded no significant difference across subjects or across song pairs. Thus, by itself, the presence of old words did not increase subjects' tendency to respond "yes, I heard this melody" when they heard a particular song.

Summary. The decrement hypothesis was not supported in the present experiment and the positive, facilitative effect of original old words on melody recognition was confirmed. Even leaving open the possibility that a

larger experiment would show a significant performance decrement when familiar-but-wrong words are present (relative to hummed conditions), we can conclude that the advantage of original old songs over mismatch songs does not depend on such a decrement in the latter condition.

Experiment 3

The purpose of Experiment 3 was to test the decrement hypothesis for text recognition rather than melody recognition. In order to conduct a rigorous test of this hypothesis and because our earlier studies had shown that recognition for our folksong texts was near ceiling, nonsense texts were used in the presentation and test sequences. Following a 24-item presentation of folksongs with nonsense texts, subjects heard a 20-item test comprising the following types of test items: (a) old songs; (b) mismatch songs; (c) old melody with new words; (d) a spoken rendition of an old nonsense text ("old words"); and (e) a spoken rendition of a new nonsense text ("new words").

The decrement hypothesis holds that text recognition is poorer in the mismatch than in the old song condition not because melody and text are integrated, but rather because the presence of a wrong melody in the mismatch condition depresses text recognition. Thus the decrement hypothesis predicts that text recognition will be poorer when an old melody is present (as measured by responses to the mismatch songs and old melody with new words) than it is when no melody is present (as measured by responses to old words and new words).

Method

Materials

We used the same set of 20 folksong pairs described previously, except that songs were sung with nonsense texts derived in the manner of Experiment 1. As much as possible, spoken texts used the rhythm of the first melody of each pair, so that spoken test items did not deviate rhythmically from the original presentation. Because of the difficulty of duplicating exact pronunciations of nonsense words, we did not record duplicate performances of all the stimuli. Thus, in the case of "old songs" and "old words" conditions, identical performances were used in the presentation and test.

Design

The design was exactly analogous to that of Experiment 2.

Procedure

The procedure was identical to that of Experiment 2, except that subjects were asked, "Did you hear this exact text before--that is, just the words?"

Subjects

Twenty Yale undergraduates with undetermined levels of musical training were paid for participating in the study. Subjects were equally divided among the five presentation/test sequences.

Results and Discussion

Responses were translated into text recognition ratings ranging from 1 to 6, as in the previous experiments. Means for the five conditions--old songs, mismatch songs, old melody with new words, old words, and new words--were 4.66, 3.73, 3.90, 2.90, and 3.23, respectively. The results of two analyses of variance were significant across subjects, $F(4,76) = 16.18$, $p < .001$, and across song pairs, $F(4,76) = 11.14$, $p < .001$.

Confirmation of the integration effect. The results were analogous to Experiment 2. Text recognition in the old song condition (mean = 4.66) was superior to that in the mismatch condition (mean = 3.73). The Scheffé procedure was significant across subjects, $p < .01$, and across song pairs, $p < .05$. This result confirms the integration effect: A nonsense text is easier to recognize when paired with its original melody than with a different, even if equally familiar melody.

Disconfirmation of the decrement hypothesis. Subjects' text recognition can be measured by difference scores (hits minus false alarms). The mean difference score when an old melody is present (mismatch minus old melody/new words) is -.18, which is not lower than -.33, the mean score when no melody is present (old words minus new words). This result fails to confirm the decrement hypothesis because the presence of a wrong melody does not depress text recognition below what it is when no melody is present. However, text recognition was so poor that old words--whether paired with a melody or not--were not rated as more familiar than new words.

On the other hand, the hypothesis that the original old melody has a positive, facilitative effect on text recognition was supported by the following results. The mean difference score when the original old melody was present (rating for old song minus rating for old melody/new words) was .76. This is significantly higher than the mean difference score when no melody was present, in the spoken condition (-.33 as above), $t(19) = -4.52$, $p < .001$ (with subjects as the sampling variable). Thus, nonsense texts were better recognized in the presence of their original old melody than on their own, in spoken form.

Criterion effects. A look at the overall means suggests that familiarity ratings were subjects to a criterion effect. Subjects were more likely to respond "yes, I heard that text" when an old melody was present (mean of mismatch and old melody/new words = 3.81) than when just the spoken text was present (mean of old words and new words = 3.06). The difference between these means is significant. (Scheffé procedure across subjects, $p < .01$, and across song pairs, $p < .01$.) Thus, the presence of a familiar melody makes the text seem more familiar, whether or not it was heard in the original presentation. This effect must be distinguished from the integration effect, which is the facilitative effect that the original melody, as opposed to a new one, has on recognition of a text that has been heard before.

General Discussion

Integration of melody and text in memory for songs is an experimental result, not an explanation, and a full account of it remains to be articulated. In the present experiments we have clarified it in two ways. First, Experiment 1 showed that the ordinary semantics of language are not required for integration. However much of the lyrics of a well-known song seem to "fit" the music, the robust effects we obtained across all of the

experiments in this and the previous article must be caused by something else. This is not to say that perhaps in ways more subtle than those evidenced here, the emotional tone of a melody could not affect subjects' interpretation of a text and hence their memory representation. But the integrative effect, at least with the present materials, does not depend on such factors.

Second, Experiments 2 and 3 showed that integration of components in song recognition is a genuine advantage of hearing the song exactly as it was before, not confusion or interference produced by a novel setting. This conclusion must be tempered by the results obtained in Experiment 2, where the decrement hypothesis was not strongly disconfirmed. Nevertheless, the advantage for "exact" old songs cannot be wholly or even primarily an artifact of interference, because positive facilitation occurred apart from this nonsignificant decrement.

By hearing the song "exactly as it was before," however, we mean the song as an abstraction rather than as an acoustic event. In Experiment 2 of the present paper and in Experiment 2 of Serafine et al. (1984), different recorded performances of the songs were used in presentation and testing. This is important in ruling out what could be called an "acoustic" hypothesis--people otherwise might recognize old songs well by seizing on some performance artifact such as a note out of tune, a vocal glitch, or even an extraneous background sound.

Clearly, melody-text integration depends neither on the acoustic identity of a re-heard song nor on semantic interaction between the components. Rather, we suggest that integration in memory may result from other, more subtle effects that melody and text have on each other. These may be thought of, broadly, as prosodic effects in that they concern the non-semantic sound pattern of either melody or text. For example, a text's consonant pattern, vowel timbres, and accents may affect the attack and decay patterns, stresses, or other aspects of tones in a melody. Consider consonant patterns. Changing "Tea for two" to "Me for you" entails changing the sound pattern from one of sudden onsets and short durations to one of gradual onsets and more prolonged durations. Such changes, even if they were to occur on melody tones that were nominally identical, would in fact change the musical quality of the tones in question. What this means is that a melody is physically different depending on the words to which it is sung. In a similar way, melody can exert an effect on the words. Patterns of pitch, loudness, stress, and articulation (e.g., staccato and legato) in a melody may affect pronunciation of individual words as well as prosody of the entire text.

If such effects were substantial, it should not be surprising that melodies are better recognized with their original words; they are in a sense "more" the same melodies than with different words or a hummed version. Likewise, a text is "more" the same words when sung to the same melody than when not.

If this reasoning is correct, then some transformation such as that used to generate nonsense words in Experiment 1 could be informative. If the mismatch conditions were constructed so that the degree of change in melody or text is minimized (by comparison to the old song) then the integration effect should be much reduced. In the example above we noted the consequences of changing "Tea for two" to "Me for you." If we changed "Gee zor goo" to "Bee vor boo" there should be much less change and correspondingly less integration.[2]

On association. We began this program of experiments out of curiosity about an unexplored point in music cognition concerning songs. Almost at once, however, we found ourselves up against fundamental issues in the ancient concept of association. We readily conceded that melody and text could become connected in the sense that presentation of one would lead to retrieval of the other. We never tested for this simple connectionism, but have no doubt our materials could be presented as paired associates and would yield, eventually, associations by this definition. Melody and text could theoretically be associated, in this sense, and yet still be represented independently. That is, each could retain its integrity as a single component and yet be attached to the other.

Our approach has insisted, at least in principle, on a different and a considerably stronger result. We require, instead, that the individual components be to some extent unrecognizable on their own, as opposed to when paired with their original companion. Thus, in this paper, we were at pains to show that the melody on its own, when hummed, was not recognized as well as when restored to its original wording; in fact recognition was close to chance. If the melody could have been recognized independently of the words, then people would have been able to do as well in the hummed condition as they did in the old song condition. This distinction between independent units attached to each other and units that undergo transformation by virtue of having been combined corresponds to the distinction between "mental compounding" and "mental chemistry" in the psychologies of William James and of John Stuart Mill, respectively (see Boring, 1957, Chapter 12).

In contemporary work on human learning and memory, our research is most closely related to Tulving's on encoding specificity (Tulving & Thomson, 1973). He, too, capitalizes on the result that when a word occurs in a particular learning context, that context can be a better aid to retrieval than the target word itself. For example, Thomson and Tulving (1970) presented the word glue as a potential learning aid next to the target word CHAIR. Later, people were better able to recall CHAIR, given the cue glue, than they were able to remember CHAIR when it was presented alone for recognition. The context apparently had changed the representation of the target (encoding specificity), just as we claim the text and melody change each other when presented together in a song. Of course, the type of change involved is quite different in songs. While Tulving's results reflect mental changes, melody and text (perhaps in addition) have physical effects on each other. What remains for future research is whether and how such changes affect the memory representation for songs.

References

Boring, E. G. (1957). A history of experimental psychology (2nd ed.). New York: Appleton Century Crofts.

Erdei, P. (1974). American folksongs to read, sing, and play. New York: Boosey and Hawkes.

Gottlieb, L. (1984). Text dependence in the young child's performance of songs. Undergraduate research project, Vassar College.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. Cognition, 16, 285-303.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. Journal of Experimental Psychology, 86, 255-262.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review, 80, 352-373.

Footnotes

[1]In our earlier studies subjects had estimated the number of songs that seemed familiar to them after a presentation of 24 excerpts from these songs, and the means of these estimates were 1.4 and 1.2, respectively, in different experiments.

[2]However, such a manipulation would also increase the tendency to confuse old and new texts, which may be an insurmountable methodological problem.

Appendix

Pairs of folksong excerpts with interchangeable texts. All folksongs from Erdei (1974).

	Number/Title	Number/Title
1.	9: Hunt the slipper	92: Cape Cod girls
2.	12: Let us chase the squirrel*	73: Christ was born*
3.	15: Who's that tapping at the window?	82: Mary had a baby
4.	16: How many miles to Babylon?*+	120: Nuts in May
5.	21: Poor little kitty puss*	80: Turn the glasses over
6.	22: Down in the meadow	68: The old woman and the pig
7.	27: Hush little baby	13: Bye, bye baby
8.	32: Bluebird	55: The old sow
9.	38: Ida Red*+	39: Mama, buy me a chiney doll
10.	52: Dear companion	88: Wayfaring stranger
11.	67: I lost the farmer's dairy key	128: Watch that lady
12.	69: Old turkey buzzard	72: My good old man
13.	78: Hold my mule	102: Needle's eye
14.	99: When the train comes along	132: Hushabye*+
15.	103: Housekeeping	147: My old hen*
16.	148: I'm going' home on a cloud	138: The raggle taggle gypsies
17.	110: Give my love to Nell*	137: Blow, boys, blow
18.	122: Cripple Creek	129: The little dappled cow
19.	142: Goodbye girls, I'm going to Boston	144: Cradle hymn
20.	2: The boatman	86: The Derby ram

*Minor alteration was made in text.
+Minor alteration was made in melody.

SOME DEVELOPMENTS IN RESEARCH ON LANGUAGE BEHAVIOR*

Michael Studdert-Kennedy†

Fifty years ago the study of language was largely a descriptive endeavor, grounded in the traditions of 19th century European philology. The object of study, as proposed by de Saussure in a famous course of lectures at the University of Geneva (1906-1911), was langue, language as a system, a cultural institution, rather than parole, language as spoken and heard by individuals. In 1933 historical linguists were describing and comparing the world's languages, tracing their family relations, and reconstructing the protolanguages from which they had sprung (Lehmann, 1973). Structural linguists were developing objective procedures for analyzing the sound patterns and syntax of a language, according to well-defined, systematic principles (e.g., Bloomfield, 1933). Students of dialect were applying such procedures to construct atlases of dialect geography (Kurath, 1939), while anthropological linguists were applying them to American Indian, African, Asian, Polynesian and many other languages (Lehmann, 1973). The work still goes on. From it we are coming to understand the origins of language diversity: not only how languages change over time and space but also how they and their dialects act as forces of social cohesion and differentiation (e.g., Labov, 1972).

However, the unfolding of the descriptive tradition and the development of new methods and theories in the field of sociolinguistics are not my concerns in this chapter. My concern, rather, is with a view of language that has emerged from a more diverse tradition. For like the taxonomic studies of Linnaeus in botany and of his followers in zoology, the great labor of language description and classification has provided the raw material for a broader science, stemming from the work of seventeenth century grammarians and of such nineteenth century figures as the German physicist Hermann von Helmholtz, the French neurologist Paul Broca, and the English phonetician Henry Sweet. The several strands that their works represent have come together over the past 30 to 40 years to form the basis of a new science of language, focusing on the individual, rather than on the social and cultural, linguistic system. Since the new focus is essentially biological, a biological analogy may be helpful. It is as though we shifted from describing and classifying the distinctive flight patterns of the world's eight or nine thousand species of birds to analyzing the basic principles of individual

---

*In N. J. Smelser & D. R. Gerstein (Eds.). (1986). Behavioral and social science: Fifty years of discovery (pp. 208-248). Washington, D.C.: National Academy Press.
†Also Queens College and Graduate Center, City University of New York.
Acknowledgment. This work was supported in part by NICHD Grant HD-01994 to Haskins Laboratories. I thank Ignatius Mattingly for comments and advice.

flight as they must be instantiated in the anatomy and physiology of every hummingbird and condor. Thus, this new science of language asks: What is language as a category of individual behavior? How does it differ from other systems of animal communication? What do individuals know when they know a language? What cognitive, perceptual and motor capacities must they have, to speak, hear, and understand a language? How do these capacities derive from their biophysical structures, that is, from human anatomy and physiology? What is the course of their ontogenetic development? And so on.

Such questions hardly fall within the province of a single discipline. The new field is markedly interdisciplinary, and addresses questions of practical application as readily as questions of pure theory or knowledge. Linguistics, anthropology, psychology, biology, neuropsychology, neurology, and communications engineering all contribute to the field, and their research has implications for workers in many areas of social import: doctors and therapists treating stroke victims, surgeons operating on the brain, applied engineers working on human-machine communication, teachers of second languages, of reading, and of the deaf and otherwise language-handicapped.

The origins of the new science are an object lesson in the interplay between basic and applied research, and between research and theory. To understand this, we must begin by briefly examining the nature of language and the properties that make it unique as a system of communication.

The Structure of Language

If we compare language with other animal communication systems, we are struck by its breadth of reference. The signals of other animals form a closed set with specific, invariant meanings (Wilson, 1975). The ultrasonic squeaks of a young lemming denote alarm; the swinging steps and lifted tail of the male baboon summon his troop to follow; the "song" of the male white-crowned sparrow informs his fellows of his species, sex, local origin, personal identity and readiness to breed or fight. Even the elaborate "dance" of the honey bee merely conveys information about the direction, distance, and quality of a nectar trove. But language can convey information about many more matters than these. In fact, it is the peculiar property of language to set no limit on the meanings it can carry.

How does language achieve this openness, or productivity? There are several key features to its design (Hockett, 1960). Here we note two. First, language is learned: it develops under the control of an open rather than a closed genetic program (Mayr, 1974). Transmission of the code from one generation to the next is therefore discontinuous: Each individual recreates the system for himself. There is ample room here for creative variation--probably a central factor in the evolution of language and in the constant processes of change that all languages undergo (e.g., Kiparsky, 1968; Locke, 1983; Slobin, 1980). One incidental consequence of this freedom is that the universal properties of language (whatever they may be) are largely masked by the surface variety of the several thousand languages, and their many dialects, now spoken in the world.

Second, and more crucially, language has two hierarchically related levels of structure. One level, that of sound pattern, permits the growth of a large lexicon; the other level, that of syntax, permits the formation of an infinitely large set of utterances. A similar combinatorial principle underlies the structure of both levels.

Consider, first, the fact that a 6-year-old, middle-class American child typically has a recognition vocabulary of some 8,000 root words, some 14,000 words in all (Templin, 1957). Most of these have been learned in the previous four years, at a rate of about five or six roots a day. As an adult, the child may come to have a vocabulary of well over 150,000 words (Seashore & Frickson, 1940). How is it possible to produce and perceive so many distinct signals?

The achievement evidently rests on the evolution in our hominid ancestors of a combinatorial principle by which a small set of meaningless elements (phonemes, or consonants and vowels) is repeatedly sampled, and the samples permuted, to form a very large set of meaningful elements (morphemes, words). Most languages have between 20 and 100 phonemes; English has about 40, depending on dialect. The phonemes themselves are formed from an even smaller set of movements, or gestures, made by jaw, lips, tongue, velum, and larynx. Thus, the combinatorial principle was a biologically unique development that provided "a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices" (Studdert-Kennedy & Lane, 1980; cf. Cooper, 1972; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). We may note, incidentally, that a large lexicon is not peculiar to complex, literate societies: Even so-called primitive human groups may deploy a considerable lexicon. For example, the Hanunoo, a stone-age people of the Philippines, have nearly three thousand words for the flora and fauna of their world (Levi-Strauss, 1966).

Of course, a large lexicon is not a language. Many languages have relatively small lexicons, and in everyday speech we may draw habitually on no more than a few thousand words (Miller, 1951). To put words to linguistic use, we must combine them in particular ways. Every language has a set of rules and devices, its syntax, for grouping words into phrases, clauses, and sentences. Among the various devices that a language may use for predicating properties of objects and events, and for specifying their relations (who does what to whom) are word order, and inflection (case, gender, and number affixes for nouns, pronouns, adjectives; person, tense, mood, and voice affixes for verbs). An important distinction is also made in all languages between open-class words with distinct meanings (nouns, verbs, adjectives, etc.) and closed-class or function words (conjunctions, articles, verbal auxiliaries, enclitics) that have no fixed meaning in themselves, but serve the purely syntactic function of indicating relations between words in a sentence or sequence of sentences. Here again then, a combinatorial principle is invoked: a finite set of rules and devices is repeatedly sampled and applied to produce an infinite set of utterances.

I should note that many of the facts about language summarily described above are already framed from the new viewpoint that has developed in the past 40 years. Let us now turn back the clock and consider the early vicissitudes of three areas of applied research that contributed to this development.

Three Areas of Applied Research in Language

In the burst of technological enthusiasm that followed World War II, federal money flowed into three related areas of language study: automatic machine translation, automatic speech recognition, and automatic reading machines for the blind. A considerable research effort was mounted in all three areas during the late 1940s and early 1950s, but surprisingly little

headway was made. The reason for this, as will become clear below, was that all three enterprises were launched under the shield of a behaviorist theory according to which complex behaviors could be properly described as chained sequences of stimuli and responses.

The initial assumption underlying attempts at machine translation was that this task entailed little more than transposing words (or morphemes) from one language into another, following a simple left-to-right sequence. If this were so, we might store a sizable lexicon of matched Russian, say, and English words in a computer and execute translation by instructing the computer to type out the English counterpart of each Russian word typed in. Unfortunately, both semantic and syntactic stumbling blocks lie in the path. The range of meanings, literal and metaphorical, that one language assigns to a word (say, English <u>high</u>, as in "high mountain," "high pitch," "high hopes," "high horse," "high-stepping," and "high on drugs") may be quite different from the range assigned by another language; and the particular meaning to be assigned will be determined by context, that is, by meanings already assigned to some, in principle, unspecifiable sequence of preceding words. Moreover, the syntactic devices for grouping words into phrases, phrases into clauses, clauses into sentences may be quite different in different languages. This is strikingly obvious when we compare a heavily inflected language, such as Russian, with a lightly inflected language with a more rigid word order, such as English. Oettinger (1972) amusingly illustrates the general difficulties with two simple sentences, immediately intelligible to an English speaker, but a source of knotty problems in both phrase structure and word meaning to a computer, programmed for left-to-right lexical assignment: <u>Time flies like an arrow</u>, and <u>Fruit flies like a banana</u>. From such observations, it gradually became clear that we would make little progress in machine translation without a deeper understanding of syntax and of its relation to meaning.

The initial assumption underlying attempts at automatic speech recognition was similar to that for machine translation and equally in error (cf. Reddy, 1975). The assumption was that the task entailed little more than specifying the invariant acoustic properties associated with each consonant and vowel, in a simple left-to-right sequence. One would then construct an acoustic filter to pass those properties but no others, and control the appropriate key on a printer by means of the output from each filter. Unfortunately, stumbling blocks lie in this path also. A large body of research has demonstrated that speech is not a simple left-to-right sequence of discrete and invariant alphabetic segments, such as we see on a printed page (e.g., Fant, 1962; Joos, 1948; Liberman et al., 1967). The reason for this, as we shall see shortly, is that we do not speak phoneme by phoneme, or even syllable by syllable. At each instant our articulators are engaged in executing patterns of movement that correspond to several neighboring phonemes, including those in neighboring syllables. The result of this shingled pattern of movement is, of course, a shingled pattern of sound. Even more extreme variation may be found when we examine the acoustic structure of the same syllable spoken with different stress or at different rates or by different speakers. From such observations it gradually became clear that we would make little progress in automatic speech recognition without a deeper understanding of how the acoustic structure of the speech signal specifies the linguistic structure of the message.

Finally, the initial assumption underlying attempts to construct a reading machine for the blind was closely related to that for automatic speech recognition and again in error (Cooper, Gaitenby, & Nye, 1984). A reading machine is a device that scans print and uses its contours to control an acoustic signal. It was supposed that, given an adequate device for optical recognition of letters on a page, one need only assign a distinctive auditory pattern to each letter, to be keyed by the optical reader and recorded on tape or played in real time to a listener--a sort of auditory Braille. Once again there were stumbling blocks, but this time they were perceptual. We normally speak and listen to English at a rate of some 150 words per minute (wpm), that is, roughly 5 to 6 syllables or 10 to 15 phonemes per second. Ten to 15 discrete sounds per second is close to the resolving power of the ear (20 elements per second merge perceptually into a low-pitched buzz). Not surprisingly, despite valiant and ingenious attempts to improve the acoustic array, even the most practiced listeners were not able to follow a substitute code at rates much beyond that of skilled Morse code receivers, namely some 10 to 15 words per minute--a rate intolerably slow for any extended use. From this work, it gradually became clear that the only acceptable output from a reading machine would be speech itself. This conclusion was one of many that spurred development of speech synthesis by artificial talking machines in following years (Cooper & Borst, 1952; Fant, 1973; Flanagan, 1983; Mattingly, 1968, 1974). The conclusion also raised theoretical questions. For example: Why can we successfully transpose speech into a visual alphabet, using another sensory modality, if we cannot successfully transpose it within its "natural" modality of sound? Why is speech so much more effective than other acoustic signals? Is there some peculiar, perhaps biologically ordained, relation between speech and the structure of language? We will return to these questions below.

I have not recounted these three failures of applied research missions to argue that money and effort spent on them were wasted. On the contrary, initial failure spurred researchers to revised efforts, and valuable progress has since been made. Reading machines for the blind, using an artificial speech output, have been developed and are already installed in large libraries (Cooper et al., 1984). There now exist automatic speech recognition devices that recognize vocabularies of roughly a thousand words, spoken in limited contexts by a few different speakers (Levinson & Liberman, 1981). Scientific texts with well-defined vocabularies can now be roughly translated by machine, then rendered into acceptable English by an informed human editor.

These advances have largely come about by virtue of brute computational force and technological ingenuity, rather than through real gains in our understanding of language. This is not because we have made no gains, for as we shall see shortly, we surely have. However, none of the devices that speak, listen, or understand actually speaks, listens, or understands according to known principles of human speech and language. For example, a speech synthesizer is the functional equivalent of a human speaker to the extent that it produces intelligible speech. But it obviously does so by quite different means than those that humans use: none of its inorganic components corresponds to the biophysical structures of larynx, tongue, velum, lips, and jaw. Instead, a synthesizer simulates speech by means of a complex system of tuned electronic circuits, and resembles a speaker somewhat as, say, a crane resembles a human lifting a weight. We are still deeply ignorant of the physiological controls by which a speaker precisely coordinates the actions of larynx, tongue, and lips to produce even a single syllable.

In short, the main scientific value of the early work I have described was to reveal the astonishing complexity of speech and language, and the inadequacy of earlier theories to account for it. One important effect of the initial failures was therefore to prepare the ground for a theoretical revolution in linguistics (and psychology) that began to take hold in the late 1950s.

The Generative Revolution in Linguistics

The publication in 1957 of Noam Chomsky's Syntactic Structures began a revolution in linguistics that has been sustained and developed by many subsequent works (e.g., Chomsky, 1965, 1972, 1975, 1980; Chomsky & Halle, 1968). To describe the course of this revolution is well beyond the scope of this chapter. However, the impact of Chomsky's writings on fields outside linguistics--philosophy, psychology, biology, for example--and their importance for the emerging science of language has been so great that some brief exposition of at least their nontechnical aspects is essential. I should emphasize that Chomsky's work has by no means gone unchallenged (e.g., Givon, 1979; Hockett, 1968; Katz, 1981). My intent in what follows is not to present a brief in its defense, but simply to sketch a bare outline of the most influential body of work in modern linguistics.

The central goal of Chomsky's work has been to formalize, with mathematical rigor and precision, the properties of a successful grammar. He defines a grammar as "a device of some sort for producing the sentences of the language under analysis" (Chomsky, 1957, p. 11). A grammar, in Chomsky's view, is not concerned either with the meaning of a sentence or with the physical structures (sounds, script, manual signs) that convey it. The grammar, or syntax, of a language is a purely formal system for arranging the words (or morphemes) of a sentence into a pattern that a native speaker would judge to be grammatically correct or at least acceptable. In Syntactic Structures, Chomsky compared three types of grammar: finite-state, phrase structure, and transformational grammars.

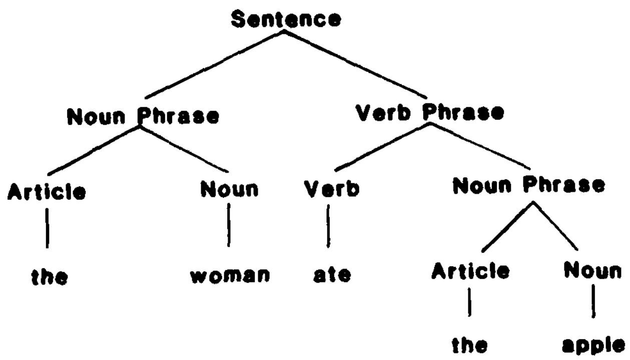
A finite-state grammar generates sentences in a left-to-right fashion: given the first word, each successive word is a function of the immediately preceding word. (Such a model is, of course, precisely that adopted by B. F. Skinner in his Verbal Behavior (1957), a dernier cri in behaviorism, published in the same year as the premier cri of the new linguistics). Chomsky (1956) proved mathematically, as work on machine translation had suggested empirically, that a simple left-to-right grammar can never suffice as the grammar of a natural language. The reason, stated nontechnically, is that there may exist dependencies between words that are not adjacent, and an indefinite number of phrases containing other nonadjacent dependencies may bracket the original pair. Thus, in the sentence, Anyone who eats the fruit is damned, anyone and is damned are interdependent. We can, in principle, continue to add bracketing interdependencies indefinitely, as in Whoever believes that anyone who eats the fruit is damned is wrong, and Whoever denies that whoever believes that anyone who eats the fruit is damned is wrong is right.

In practice, we seldom construct such sentences. However, the recursive principle that they illustrate is crucial to every language. The principle permits us to extend our communicative reach by embedding one sentence within another. For example, even a four-year-old child may combine, We picked an

apple and I want an apple for supper into the utterance, I want the apple we picked for supper. Thus, the child embeds an adjectival phrase, we picked (= that we picked with the relative pronoun deleted), to capture two related sentences in a single utterance (cf. Limber, 1973).

Chomsky goes on to consider how we might formulate an alternative and more powerful grammar, based on the traditional constituent analysis of sentences into "parts of speech." Constituent analysis takes advantage of the fact that the words of any language (or an equivalent set of words and affixes) can be grouped into categories (such as noun, pronoun, verb, adjective, adverb, preposition, conjunction, article) and that only certain sequences of these categories form acceptable phrases, clauses, and sentences. By grouping grammatical categories into permissible sequences, we can arrive at what Chomsky terms a phrase-structure grammar. Such a grammar is "a finite set...of initial strings and a finite set...of 'instruction formulas' of the form X→Y interpreted: 'rewrite X as Y'" (Chomsky, 1957, p. 29). Figure 1 illustrates a standard parsing diagram of the utterance, The woman ate the apple, in a form familiar to us from grammar school (above), and as a set of "rewrite rules" from which the parsing diagram can be generated (below).

Parsing Diagram



Rewrite Rules

(1) Sentence ⟶ Noun Phrase + Verb Phrase
(2) Noun Phrase ⟶ Article + Noun
(3) Verb Phrase ⟶ Verb + Noun Phrase
(4) Article ⟶ {the, a}
(5) Noun ⟶ {woman, apple...}
(6) Verb ⟶ {ate, seized...}

Figure 1. Above, a parsing diagram dividing the sentence The woman ate the apple into its constituents. Below, a set of rewrite rules that will generate any sentence having the constituent structure shown above.

Notice, incidentally, that rewrite rules are indifferent to meaning. They will generate anomalous utterances such as The chocolate loved the clock, no less readily than The woman ate the apple. Moreover, many native speakers would be willing to accept such anomalous utterances as grammatically correct, even though they have no meaning. This hints at the possibility that syntactic capacity might be autonomous, a relatively independent component of the language faculty. This is a matter to which we will return below.

An important point about a set of rewrite rules is that it specifies the grouping of words necessary to correct understanding of a sentence. The sentence Let's have some good bread and wine is ambiguous until we know whether the adjective good modifies only bread or both bread and wine. The distinction may seem trivial. But, in fact, the example shows that we are sensitive (or can be made sensitive) to an ambiguity that could not have arisen from any difference in the words themselves or in their sequence. Rather, the origin of the ambiguity lies in our uncertainty as to how the words should be grouped, that is, as to their phrase structure. A correct (or incorrect) interpretation of their meaning therefore depends on the listener (and a fortiori the speaker) being able to assign an abstract phrase structure to the sequence of words.

Whether a complete grammar of English, or any other natural language, could be written as a set of phrase-structure rules is not clear. In any event, Chomsky argues in Syntactic Structures that such a grammar would be unnecessarily repetitive and complex, since it does not capture a native speaker's intuition that certain classes of sentence are structurally related. For example, the active sentence Eve ate the apple and the passive sentence, The apple was eaten by Eve could both be generated by an appropriate set of phrase-structure rules, but the rules would be different for active sentences than for their passive counterparts. Surely, the argument runs, it would be "simpler" if the grammar somehow acknowledged their structural relation by deriving both sentences from a common underlying "deep structure." The derivation would be accomplished by a series of steps or "transformations" whose functions are to delete, modify, or change the order of the base constituents Eve, ate, apple.

An important aspect of transformations is that they are structure dependent, that is, they depend on the analysis of a sentence into its structural components, or constituents. For example, to transform such a declarative sentence as The man is in the garden into its associated interrogative Is the man in the garden?, a simple left-to-right rule would be: "Move the first occurrence of is to the front." However, the rule would not then serve for such a sentence as The man who is tall is in the garden, since it would yield Is the man who tall is in the garden? The rule must therefore be something like: "Find the first occurrence of is following the first noun phrase, and move it to the front" (Chomsky, 1975, pp. 30-31). Thus, a transformational grammar, no less than a phrase-structure grammar, presupposes analysis of an utterance into its grammatical (or phrasal) constituents. We may note, in passing, that children learning a language never produce sentences such as Is the man who tall is in the garden? Rather, their errors suggest that, even in their earliest attempts to frame a complex sentence, they draw on a capacity to recognize the structural components of an utterance.

However, here we should be cautious. Chomsky has repeatedly emphasized that "...a generative grammar is not a model for a speaker or hearer" (1965, p. 9), not a model of psychological processes presumed to be going on as we speak and listen. The word "generative" is perhaps misleading in this regard. Certainly, experimental psychologists during the 1960s devoted much ingenuity and effort to testing the psychological reality of transformations (for reviews, see Cairns & Cairns, 1976; Fodor, Bever, & Garrett, 1974; Foss & Hakes, 1978). But the net outcome of this work was to demonstrate the force of Chomsky's distinction between formal descriptions of a language and the strategies that speakers and listeners deploy in communicating with each other (cf. Bever, 1970).

At first glance, the distinction might seem to be precisely that between _langue_ and _parole_, drawn by de Saussure. However, for de Saussure, _langue_, the system of language, "exists only by virtue of a sort of contract signed by the members of a community" (de Saussure, 1966, p. 14): it is a kind of formal artifice or convention, maintained by social processes of which individuals may be quite unaware. By contrast, for Chomsky the "generative grammar [of a language] attempts to specify what the speaker actually knows" (1965, p. 8). What a speaker knows, _competence_ in Chomsky's terminology, is attested to by "intuitive" judgments of grammaticality. What a speaker does, _performance_ (_parole_), is linguistic competence filtered through the indecisions, memory lapses, false starts, stammerings, and the "thousand natural [nonlinguistic] shocks that flesh is heir to." Thus, even though a theory of grammar is not a theory of psychological process, it _is_ a theory of individual linguistic capacity.

In Chomsky's view, the task of linguistics is to describe the structure of language much as an anatomist might describe the structure of the human hand. The complementary role of psychology in language research is to describe language function and its course of behavioral development in the individual, while physiology, neurology, and psychoneurology chart its underlying structures and mechanisms.

Whether this sharp distinction between language as a formal object and language as a mode of biological function can, or should, be maintained is an open question. What is clear, however, is that it was from a rigorous analysis of the formal properties of syntax (and, later, of phonology: see Chomsky & Halle, 1968) that Chomsky was led to view language as an autonomous system, distinct from other cognitive systems of the human mind (cf. Fodor, 1982; Pylyshyn, 1980). His writings during the late 1950s and 1960s brought an exhilarating breath of fresh air to psychologists interested in language, because they offered an escape from the stifling behavioristic impasse, already noted by Lashley (1951) and others (e.g., Miller, Galanter, & Pribram, 1960).

The result was an explosion of research in the psychology of language, with a strong emphasis on its biological underpinnings. Whatever one's view of generative grammar, it is fair to say that almost every area of language study over the past 25 years has been touched, directly or indirectly, whether into action or into reaction, by Chomsky's work. This will be obvious from the following selective review of research in four major areas: acoustic phonetics, American Sign Language (ASL), brain specialization for language, and language development in children.

Acoustic Phonetics

We begin with audible speech, partly because we are then following the course of development, both in the species and the individual, from the bottom up; partly because it is in this area, where we are dealing with observable, physical processes, that the most dramatic progress has been made; and partly because we have come to realize in recent years that the physical medium of language places fundamental constraints on its surface structure. To understand this we must know something of the way speech is produced.

The source-filter theory of speech production. The source-filter theory, first proposed by Johannes Müller in 1848, has been elaborated in the past 50 years, notably at the University of Tokyo (Chiba & Kajiyama, 1941), the Royal Institute of Technology in Stockholm (Fant, 1960, 1973) and, in this country, the Massachusetts Institute of Technology (Stevens & House, 1955, 1961) and Bell Telephone Laboratories (Flanagan, 1983). As a result of this work, we are now able to specify accurately the possible acoustic outputs of any vocal tract, animal or human.

When we speak, we drive air from our lungs through the pharynx, mouth, teeth, lips and, sometimes, nose. The sound source is usually either the "voice" produced by rapid pulsing of the vocal cords (as in the final sounds of be and do), the hiss of air blown through a narrow constriction (as in the initial and final sounds of safe and thrush) or both (as in the final sounds of leave and bees). The resonant filter is the vocal tract, its air set into vibration by the flow of air from the lungs, much as we produce sound from a bottle or a wind instrument by blowing air across its top.

To some large degree linguistic information (that is, consonants and vowels) is conveyed by systematic variations in the configuration of the vocal tract. For example, if we lower the tongue and move it back toward the pharynx, we set up a pattern of resonances (known as formants) corresponding to the vowel [a]. If we raise the tongue forward toward the gums, we set up resonances for the vowel [i]. Finally, if we raise the tongue backward toward the soft palate, we set up resonances for the vowel [u]. These three sounds are the most distinct vowels, both articulatorily and acoustically, that the human vocal tract can produce, and all known languages use at least two of them.

(We may note, in passing, that Lieberman and his colleagues [Lieberman & Crelin, 1971; Lieberman et al., 1972]) have used the source-filter theory of speech production to demonstrate that these vowels lie outside the range of sounds that could be produced either by an adult chimpanzee or by a newborn human infant. The reason for this is that the larynx in both chimpanzee and infant is high in the throat, restricting the range of possible tongue movements. An advantage of the high larynx for the infant is that it provides an arrangement of the oral tract such that, like other mammals, the infant can suck through its mouth and breathe through its nose at the same time. Over the first six months of life, the infant's larynx lowers, a special swallowing reflex develops to prevent food entering the lungs, and the infant becomes capable of producing the vowels of the language spoken around it. The lowered larynx seems to be one of several adaptations of the vocal apparatus that have suited it for speaking as well as for eating and breathing.)

Of course, we do not speak only in vowels. Rather, we speak in runs of syllables, alternately constricting the vocal tract to form consonants, opening it to form vowels. (This repeated opening and closing of the tract produces the rises and falls of amplitude that are the basis of speech rhythm and poetic meter.) What is of interest, as we have already remarked, is that the tract configurations appropriate to particular consonants and vowels do not follow each other in linear sequence. At any instant, each articulator is executing a complex pattern of movement, of which the spatiotemporal coordinates reflect the influence of several neighboring segments. Readers may test this by slowly uttering, for example, the words <u>cool</u> and <u>keel</u>. They will find that the position of the tongue on the palate during closure for the initial consonant, [k], is slightly further back for the first word than for the second. The result of this interleaving is that, at any instant, the sound is conveying information about more than one phonetic segment, and that each phonetic segment draws information from more than one piece of sound--an obvious problem for automated speech recognition. Unfortunately, we cannot, as was at one time hoped, escape from this predicament by building a machine to recognize syllables, because similar interactions between phonetic segments occur across syllable boundaries. We see all this quite clearly if we examine a sound spectrogram.

<u>The sound spectrograph</u>. The sound spectrograph was developed at Bell Telephone Laboratories during World War II, to provide a visible display of the acoustic spectrum of speech as it changes over time. Originally, it was hoped that the device would enable deaf persons to use the telephone (Potter, Kopp, & Green, 1947), but this proved impracticable because spectrograms are formidably difficult to read (though see Cole et al., 1980).

Figure 2 is a spectrogram of the utterance <u>She began to read her book</u>. Frequency on the ordinate is plotted against time on the abscissa. Variations in relative amplitude appear as variations in the darkness of the pattern. The dark bars correspond to formants, that is, to resonant peaks in the vocal tract resonance function. Scattered patches, as at the beginning, correspond to the noise of fricatives, e.g., [f], [s], and stop consonants, e.g., [p], [b]. A series of vertical lines has been drawn, dividing the spectrogram into discrete, acoustic segments. There are 25 of these segments, even though the utterance consists of only 17 phonetic segments and 7 syllables. Some of these acoustic segments correspond more or less directly to phonetic segments: thus, segments 1 and 2 correspond to the two sounds of <u>she</u>. Segment 3, on the other hand, corresponds to the first three sounds of <u>began</u>, segments 11 and 12 to the first sound of <u>to</u>, segment 23 to the first two sounds of <u>book</u>.

The sound spectrograph revealed, for the first time, the astonishing variability of the speech signal both within and across speakers. It was also the basis for the first systematic studies of speech perception, from which we have learned which aspects of the signal carry crucial phonetic information. These studies, in turn, provided the basis for the development of speech synthesis. Thus, artificial talking machines, now being used in reading machines for the blind and in a variety of human-machine communication systems, rest squarely on the shoulders of the spectrograph.

<u>Speech perception</u>. Early work in speech perception was largely guided by the demands of telephonic communication. Its aim was to estimate how much distortion (by filtering, noise, peak-clipping, and so on) could be imposed on the signal without seriously reducing its intelligibility (Licklider & Miller,

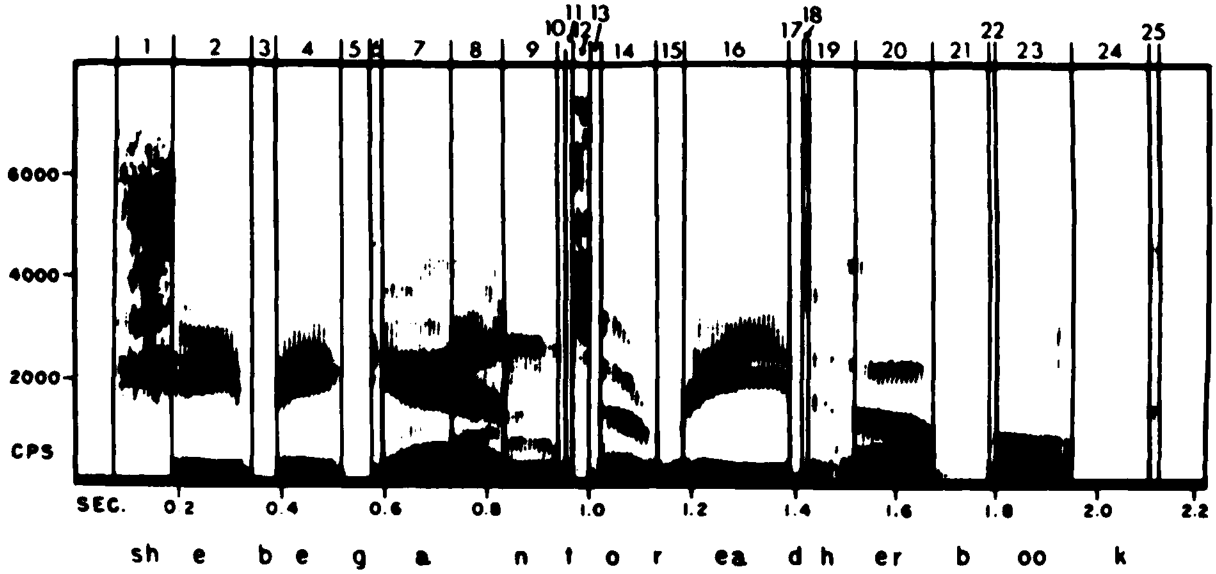


Figure 2. A spectrogram of the utterance She begain to read her book. Frequency is plotted on the ordinate, time on the abscissa; relative amplitude is represented by varying degrees of darkness in the display. The dark horizontal bands reflect resonant peaks in the vocal tract transfer function (formants, conventinally numbered from the bottom up: first formant, second formant, etc.); the vertical striations reflect repeated opening and closing of the glottis (voice). Heavy vertical lines have been drawn dividing the pattern into 25 discrete acoustic segments (see text).

1951; Miller, 1951). Two general conclusions from this work were surprising and important. First, speech is so resistant to distortion that we can throw away large parts of the signal without reducing its intelligibility. Second, intelligibility does not depend on naturalness. These two facts made it possible to learn a great deal about the important information-bearing elements in speech by stripping it down to its minimal cues.

Work of this kind was first undertaken at Haskins Laboratories in New York during the 1950s, as part of a program to develop a suitable output for a reading machine. The key research tool was the Pattern Playback, developed by F. S. Cooper (Cooper, 1950; Cooper & Borst, 1952) to reconvert the visual pattern of a spectrogram into sound. The pattern, painted on a moving acetate belt, reflects frequency-modulated light to a photocell that drives a speaker. Figure 3 illustrates an early spectrogram and its stylized copy. If the copy is passed through the playback, it produces an intelligible version of the utterance To catch pink salmon. The utterance sounds unnatural, partly because the formant bandwidths have been sharply reduced, partly because it is spoken in a monotone.

The playback made it possible for experimenters to manipulate the speech signal systematically, by pruning, deleting, or exaggerating portions of the spectrographic pattern until they had determined the minimal cues for any

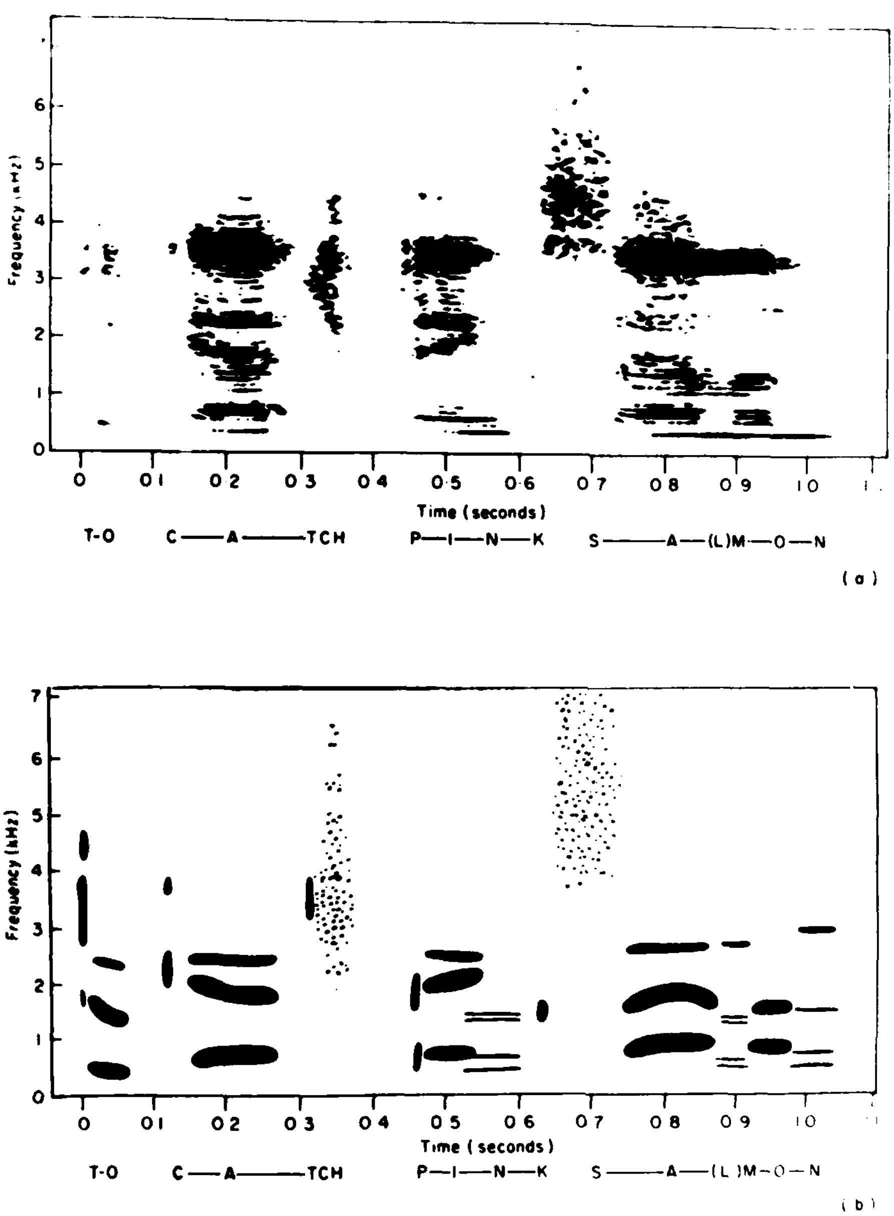


Figure 3. Above, a spectrogram of the utterance <u>To catch pink salmon</u>. Below, a stylized copy of the spectrogram, sufficient to regenerate the utterance if played on the Pattern Playback.

particular utterance (Liberman, 1957; Liberman et al., 1959). With this device, and with its successors at Haskins and elsewhere, a body of knowledge was built up, sufficient for synthesis by rule of relatively high-quality speech (Fant, 1960, 1968; Flanagan, 1983; Mattingly, 1974).

Several reviews of the perceptual implications of this work have been published (Darwin, 1976; Liberman et al., 1967; Liberman & Studdert-Kennedy, 1978; Studdert-Kennedy, 1974, 1976), and I will not review them here. However, two facts deserve note. First, the cues for a given phonetic segment (that is, for a particular consonant or vowel) vary markedly as a function of context. Figure 4 displays spectrograms of the naturally spoken syllables [did] and [dud]. We know from synthetic speech that a main cue to the initial [d] lies in changes in the second formant after onset. Notice that the second formant rises before [i], falls before [u], and that the rising and falling patterns are precisely reversed for the final [d]. Yet all are heard as [d]. Moreover, if these patterns or their synthetic versions are removed from context and presented to listeners for judgments, they are no longer heard as [d], nor are they heard as invariant. Rather they are heard as rising and falling tones (Liberman et al., 1967). In other words, different acoustic patterns are heard as different in a nonspeech context but as the same in a speech context. This is merely one of dozens of such examples.

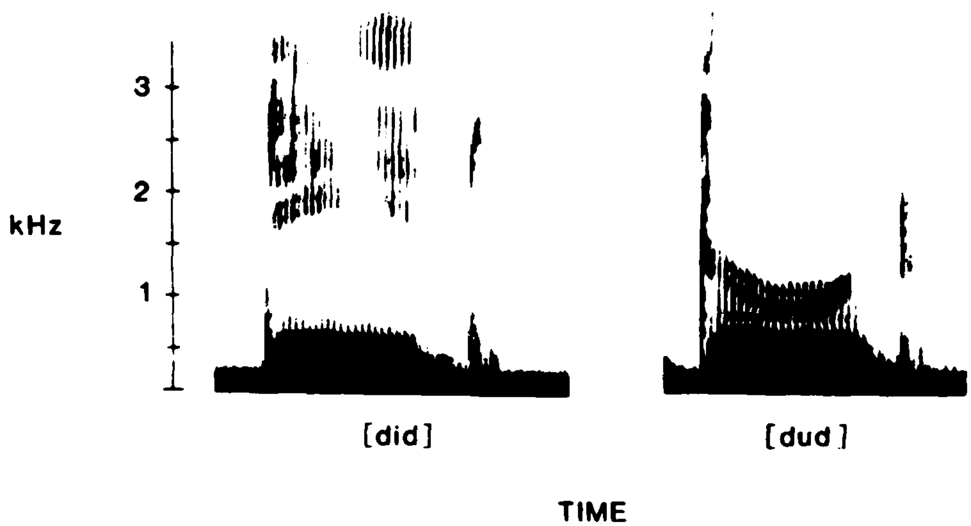


Figure 4. Spectrograms of naturally spoken [did] (deed) and [dud] (dood). The acoustic information specifying the alveolar place of articulation of the initial and final consonants is primarily carried by the second formant, centered around 2kHz for [did] and slightly below 1 kHz for [dud]. Note that this formant forms a parabola, concave downwards in [did], concave upwards in [dud]. Despite this difference, both patterns are heard as beginning and ending with [d].

The second fact of note is that despite the apparent lack of discrete phonetic segments in the signal, listeners have little difficulty in learning to find segments--so little, in fact, that a segmental representation of speech is the basis of the alphabet.

The interpretation of these facts is still a matter of controversy (e.g., Cole & Scott, 1974; Ladefoged, 1980; Stevens, 1975), and I will not pursue the matter here. However, it is worth noting that such findings gave rise to the hypothesis that humans have evolved a specialized perceptual mechanism for speech, distinct from, though dependent on, their general auditory system (Liberman, 1970, 1982; Liberman et al., 1967; Liberman & Studdert-Kennedy, 1978). The hypothesis has received substantial support from many dozens of studies of dichotic listening over the past 20 years (e.g., Kimura, 1961, 1967; Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970; for a review, see Porter & Hughes, 1983). The conclusion from this work, and from studies of patients with separated cerebral hemispheres (see section below on brain specialization for language), is that the left hemisphere of most normal right-handed individuals is specialized not only for speaking (as has been known for many years from studies of brain-damaged patients), but also for perceiving speech. Specifically, there is now good reason to believe that "while the general auditory system common to both hemispheres is equipped to extract the auditory parameters of a speech signal, the dominant [i.e., left] hemisphere may be specialized for the extraction of linguistic features from these parameters" (Studdert-Kennedy & Shankweiler, 1970, p. 579).

An important implication of this conclusion is that speech forms an integral part of the left-hemisphere language system discussed below. With this in mind let us turn to recent work on American Sign Language, which draws on a different perceptuomotor system than spoken language.

American Sign Language

Speech is the natural medium of language. Specialized structures and functions have evolved for spoken communication: vocal tract morphology, lip, jaw, and tongue innervation, mechanisms of breath control (Lenneberg, 1967), and perhaps even (as I have just suggested) matching perceptual mechanisms. But is there any further specialization for language? Is language an autonomous system, distinct from other cognitive systems, as Chomsky has argued?

An opportunity to address this question has arisen in recent years from an unexpected quarter: sign languages of the deaf. Until some 20 years ago, it was commonly believed that sign languages of the deaf--and of other social groups, such as American Plains Indians and Australian aborigines--were either more or less impoverished hybrids of conventional iconic gesture and impromptu pantomime, or artificial systems based, like reading and writing, on a specific spoken language. Artificial systems, such as Signed English and Paget-Gorman, are indeed used in many schools of the deaf: their signs refer to letters (finger-spelling) or higher-order linguistic units (words, morphemes), and their syntax follows that of the base language. However, there are other signed languages, not based on any spoken language, with their own independent lexicons and syntactic systems. The most extensively studied of these is American Sign Language (ASL), the first language of over 100,000 deaf individuals and, according to Mayberry (1978), the fourth most common language (after English, Spanish, and Italian) in the United States.

Modern ASL stems from a French-based sign language introduced into the United States by Thomas Gallaudet in 1817. (According to Stokoe [1974] ASL signers today find French SL more intelligible than British SL, a nice demonstration that ASL is independent of English.) Thus, the original language was in fact based on a spoken language. However, over the past 165 years it has developed among the deaf into an independent sign language.

Structural analysis of ASL was first undertaken by Stokoe (1960), and in 1965 he and his colleagues (Stokoe, Casterline, & Croneberg, 1965) published <u>A Dictionary of American Sign Language on Linguistic Principles</u>, containing a description and English gloss of nearly 2500 signs. The dictionary used minimal pair analysis to show that signs contrasted along three independent dimensions: hand configuration, place of articulation, and movement. For example, signs for APPLE and JEALOUS contrast in hand configuration; signs for SUMMER and UGLY contrast in place of articulation; signs for CHAIR and TRAIN contrast in movement (Klima & Bellugi, 1979, p. 42). Stokoe et al. isolated 55 "cheremes" or primes, analogous to the phonemes of a spoken language: 19 for hand configuration, 12 for place of articulation, and 24 for movement. Thus, they demonstrated that ASL has a sublexical structure, analogous to the phonological structure of a spoken language.

ASL also has a second level of structure, a grammar or syntax. This has been demonstrated in an extensive program of research at the Salk Institute for Biological Studies in La Jolla, over the past 10 years (Klima & Bellugi, 1979). I will not attempt to review this work in any detail, but several points deserve note. First, ASL has a rule-governed system of compounding, by which signs may be combined to form a new sign different in meaning from its components. The process is analogous to that by which, in English, <u>hard</u> and <u>hat</u>, say, are combined to form <u>hardhat</u>, meaning a construction worker. Thus, the lexicon of ASL can be expanded by rule, not simply by iconic invention.

Second, ASL has an elaborate system of inflections by which it modulates the meaning of a word. For example, in English, changes in aspectual meaning (that is, distinctions in the onset, duration, frequency, recurrence, permanence, or intensity of an event) are indicated by concatenating morphemes. We may say, <u>he is quiet</u>, <u>he became quiet</u>, <u>he used to be quiet</u>, <u>he tends to be quiet</u>, and so on. All these meanings are conveyed in ASL by distinct modulations of the root sign's movement. In the root sign for QUIET the hands move straight down from the mouth, while for TENDS TO BE QUIET they move down forming a circle. Similarly, related nouns and verbs are also distinguished by movements, while verbs are inflected by movement modulation for person, number, reciprocal action, and aspect.

Third, ASL has a spatial (rather than a temporal) syntax. Nouns introduced into a discourse are assigned arbitrary reference points in a horizontal plane in front of the signer. These points then serve to index grammatical relations among referents: verb signs are executed with a movement between two points, or across several points, to indicate subject and object. Thus, a grammatical function variously served in spoken language by word order, case markers, verb inflections, and pronouns is fulfilled in ASL by a spatial device.

Finally, ASL has a variety of syntactic devices that make use of the face. Liddell (1978) has shown that a relative clause ("The apple <u>that Eve offered</u> tempted him") may be marked by tilting back the head, raising the

eyebrows, and tensing the upper lip for the duration of the clause. Baker and Padden (1978) describe gestures of the face and head that mark the juncture of conditional clauses ("If you eat the fruit, you will be punished").

In short, though structural analysis of ASL is far from complete, it is evident that the language has a dual pattern of form and syntax, fully analogous to that of a spoken language. Nonetheless, there are differences. The main structural difference between ASL and English was illustrated by Klima and Bellugi (1979) in a comparison of their rates of communication. The times taken to tell a story in the two languages were almost exactly equal. Yet the speaker used two to three times as many words as the signer used signs. The reason for the discrepancy, already hinted at, lies in the temporal distribution of information. Speech, for the most part, develops its patterns in time, sequentially, while ASL develops its patterns both simultaneously, in space, and sequentially. The difference is evidently due to the difference in the perceptual modalities addressed. Sign, addressed to the eye, is free to package information in parallel; speech, addressed to the ear, is forced into a serial mode. What is interesting, of course, is that despite constraints of modality, the two languages convey information at roughly the same rate. This suggests that they may be operating under the same temporal constraints of cognition.

What, finally, are the implications of this work for the study of speech and language? Evidently, the dual structure of language is not a mere consequence of perceptuomotor modality, but a reflection of cognitive requirements. Whether these cognitive requirements are linguistic rather than general is still not clear. Differently put, we still do not know whether the relation between signed and spoken language is one of analogy or homology. If the two systems prove to be homologous, that is, if they prove to draw on the same neural structures and organization, we will have strong evidence that language is a distinct cognitive faculty. However, if they do not draw on the same underlying neural organization, we might suppose that linguistic structure is purely functional, the adventitious consequence of a cognitively complex animal's attempt to communicate its thought. Studies of sign-language breakdown due to brain injury, discussed below, are therefore of unusual interest and importance.

Brain Specialization for Language

Most of our knowledge of brain specialization for language comes from those "experiments of nature" in which some more or less circumscribed lesion (due to stroke, epilepsy, congenital malformation, gunshot wounds, and so on) proves to be correlated with some more or less circumscribed cognitive or linguistic deficit (for a brief account of modern brain-scanning techniques, see Benson, 1983, and references therein). Recently, our sources of knowledge have been expanded by use of brain stimulation, preparatory to surgery under local anesthesia (Ojemann, 1983, and references therein), and by studies of so-called "split-brain" patients whose cerebral hemispheres have been separated surgically for relief of epilepsy (see below). Some degree of concordance between patterns of brain localization in normal and abnormal individuals has been established by experiments on normals in which visual or auditory input is confined, or more clearly delivered, to one hemisphere rather than the other (Moscovitch, 1983).

<u>Evidence from studies of aphasia</u>. The term <u>aphasia</u> refers to some impairment in language function, whether of comprehension, production, or both, due to some more or less well-localized damage to the brain. Systematic study of aphasia goes back well over a hundred years, and the literature of the subject is vast (for reviews, see, for example, Goodglass & Geschwind, 1976; Hecaen & Albert, 1978; Lesser, 1978; Luria, 1966, 1970). The most that can be done here is to hint at one area in which linguistics (that is, formal language description) has begun to affect aphasia studies.

Until recently, the standard framework for describing aphasic symptoms was that of the language modalities: speaking, listening, reading, and writing, or, more generally, the dimensions of expression and reception. These are still the dimensions of the major test batteries used to diagnose aphasia, such as the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1972). An important assumption, underlying any attempt at diagnosis, is that damage to a particular region of the brain has particular, not general, effects on language function. The assumption has strong empirical support and has led to the isolation of two (among several other) broad types of aphasia, nonfluent and fluent, respectively associated with damage to the left cerebral hemisphere in an anterior region around the third frontal convolution (Broca's area) and a posterior region around the superior temporal convolution (Wernicke's area).

Broca's area lies close to the motor strip of the cortex (in fact, close to that portion of the strip associated with motor control of the jaw, lips, and tongue), while Wernicke's area surrounds the primary auditory region. In accord with this anatomical dissociation, a Broca's aphasic (that is, an individual with damage to Broca's area) has been classically found to be nonfluent: having good comprehension, but awkward speech, characterized by pauses, difficulties in word-finding, and distorted articulation; utterances are described as "telegrammatic," consisting of simple, declarative sentences, relying on nouns and uninflected verbs, omitting grammatical morphemes or function words. By contrast, a Wernicke's aphasic has been found to have poor comprehension, even of single words, but fluent speech, composed of inappropriate or nonexistent (though phonologically correct) words, often inappropriately inflected and/or out of order.

Notice that these descriptions are still couched in terms of input and output--that is, modalities of behavior--rather than in linguistic terms. The idea that linguistic theory should be brought to bear on aphasia, and attempts made to characterize deficits in terms of overarching linguistic function, has been proposed a number of times in the past (e.g., Jakobson, 1941; Pick, 1913). But only recently (again, partly under the influence of Chomsky's view of language as an autonomous system, composed of autonomous syntactic and phonological subsystems) has the idea begun to receive widespread attention. The general hypothesis of the studies described below is that language breaks down along linguistic rather than modal lines of demarcation.

We will focus mainly on the hypothesis that syntactic competence is discretely and coherently represented in Broca's area of the left frontal lobe. If this is so, the clinical impression that Broca's aphasics have good comprehension, despite their agrammatic speech (and, incidentally, writing), must be in error. More careful testing should reveal deficits in their comprehension, also.

Caramazza and Zurif (1976) tested this hypothesis with three types of sentence: (1) Simple declarative sentences in which semantic constraints might permit decoding without appeal to syntax (The apple that the boy is eating is red); (2) so-called reversible sentences that require knowledge of syntactic relations for decoding (The boy that the girl is chasing is tall); and (3) implausible, though grammatically correct, sentences (The boy that the dog is patting is fat). The sentences were presented orally and patients were asked to choose which of two pictures represented the meaning of the sentence. The incorrect alternative showed either a subject-object reversal or an action different from that specified by the verb.

Broca's aphasics performed very well on simple declarative sentences and on sentences with strong semantic constraints (as when the incorrect alternative depicted the wrong action). On reversible plausible and implausible sentences (when the incorrect alternative depicted a subject-object reversal) the patients' performance was at chance. Caramazza and Zurif (1976) concluded that the clinical impression of good comprehension in Broca's aphasics was due to their ability to draw on semantic and pragmatic constraints to understand sentences despite their inability to process syntax.

Other studies have shown that Broca's aphasics a) have difficulty in parsing a sentence into its grammatical constituents (von Stockert, 1972); b) cannot use articles to assign appropriate reference in understanding a sentence (Goodenough, Zurif & Weintraub, 1977), and c) cannot, in general, access closed-class grammatical morphemes (Zurif & Blumstein, 1978). These studies are not without their critics (e.g., Linebarger, Schwartz, & Saffran, 1983), nor is the general claim that aphasic breakdown is typically (or, indeed, ever) along purely linguistic lines (Studdert-Kennedy, 1983, pp. 193-194): the locus and extent of brain damage in aphasia is largely a matter of chance, and it is rare that language alone is affected. However, we have other sources of evidence to test the hypothesis that syntax is represented in the brain as a functionally discrete subsystem.

Evidence from split-brain studies. One source of evidence is the split-brain patient whose cerebral hemispheres have been separated surgically for relief of epilepsy. The condition permits an investigator to assess the cognitive and linguistic capacities of each hemisphere separately. Zaidel (1978) has devised a contact lens, opaque on either the nasal or temporal side, that can be used (profiting from decussation of the optic pathways) to ensure that visual information is freely scanned by a single hemisphere. A variety of written verbal materials--nonsense syllables, words, sentences of varying length and complexity--and pictures can then be used to test the capacities of the isolated hemispheres. For example, the sentences, The fish is eating or The fish are eating, can be presented to a single hemisphere, together with appropriate alternative pictures, to test the hemisphere's capacity to understand written verbal auxiliaries (is, are) (Zaidel, 1983). Similarly, pictures of various objects belonging to different classes (fruit, furniture, vehicles, etc.) might be presented to a single hemisphere to test the hemisphere's capacity to categorize.

The number of available subjects is, of course, limited. But the conclusions from studies of four split-brain patients are remarkably consistent (Zaidel, 1978, 1980, 1983). In general, each hemisphere seems to have "a complete cognitive system with its own perception, memory, language, and cognitive abilities, but with a unique profile of competencies: good on some abilities, poor on others" (Zaidel, 1980, p. 318). Of particular

interest in the present context is the finding that, although the right hemisphere cannot speak, it has a sizable auditory and reading lexicon. However, unlike the left hemisphere, the right cannot read new (nonsense or unknown) words or recognize words for which it has no semantic interpretation. Similarly, the right hemisphere cannot group pictures of objects on the basis of rhyme (e.g., nail, male). Evidently, phonological analysis is the prerogative of the left hemisphere.

The syntactic capacity of the right hemisphere is also limited. The hemisphere can recognize verbal auxiliaries (see above), but has difficulty in discriminating inflections (The fish eat vs. The fish eats). Similarly, the right hemisphere can recognize and interpret nouns, adjectives, and certain prepositions, but has difficulty with the English infinitive marker to. These findings on closed-class morphemes mesh to a degree with the deficits of Broca's aphasics, described above. Not surprisingly, the right hemisphere's capacity to understand sentences is sharply reduced: it cannot deal with sentences longer than about three words.

On the evidence of these studies, then, the right hemisphere has essentially no phonological capacity and only a limited syntactic capacity. Unfortunately, the limited syntactic capacity is equivocal because all these split-brain patients have had epilepsy since early childhood. Brain disorders are known to lead to reorganization and redistribution of function, particularly in childhood (Lenneberg, 1967; Dennis, 1983). We cannot therefore be sure that such syntactic capacity as the right hemisphere displays does not reflect compensation for left hemisphere deficiencies, induced by epilepsy.

Evidence from studies of ASL "aphasia." Studies of normally hearing, brain-damaged patients have established a double dissociation of brain locus and function in right-handed individuals: the left cerebral hemisphere is specialized for language, the right hemisphere for visual-spatial functions (as revealed, for example, by tests requiring a subject to copy a drawing, assemble wooden blocks into a pattern, or discriminate between photographs of unfamiliar faces). As we have seen, ASL is an autonomous linguistic system with a dual structure analogous to that of spoken language, on the one hand, yet, on the other, it encodes its meanings in visual-spatial rather than auditory-temporal patterns. How then should we expect brain damage to affect the language of a native ASL signer?

The answer bears directly on our understanding of the basis of brain specialization for language. For if language loss in ASL aphasia follows damage to the right hemisphere, we may infer that language is drawn to the hemisphere controlling its perceptuomotor channel of communication. But if language loss follows damage to the left hemisphere, we may infer that the neural structure of that hemisphere is, in some sense, matched to the structure of language, whatever its modality. Language might then be seen as a distinct cognitive faculty, sufficiently abstract in its descriptive predicates to encompass both speaking and signing.

Recent studies at the Salk Institute, the first systematic and linguistically motivated studies of ASL aphasia on record, support the second hypothesis. Moreover, the forms of ASL breakdown vary with locus of lesion in a fashion strikingly similar to certain forms of spoken-language breakdown. Bellugi, Poizner, and Klima (1983) describe three patients, all of whom are native ASL signers and display normal visual-spatial capacity for nonlanguage

functions. Their symptoms, resulting from strokes, divide readily into the two broad classes noted above for spoken language: two patients are fluent, one is nonfluent.

The two fluent patients display quite different symptoms, coordinated with different areas of damage to the left hemisphere. The deficits of one patient (PD) are primarily grammatical; the deficits of the other (KL) are primarily lexical. PD has extensive subcortical damage from below Broca's area in the frontal lobe through the parietal to the temporal lobe, abutting Wernicke's area. PD produces basically normal root signs, but displays an abundance of semantic and grammatical paraphasias. He produces many semantically displaced signs (e.g., EARTH for ROOM, BED for CHAIR, DAUGHTER for WIFE). More strikingly, he often modulates an appropriate root form with an inappropriate or nonsensical inflection. Finally (despite his normal, nonlanguage visual-spatial capacity), his spatial syntax is severely disordered: he misuses or avoids spatial indexing (the equivalent of pronominal function, as noted above), and overuses nouns.

The second fluent patient, KL, has more limited damage, extending in a strip across the left parietal lobe. Her deficits, though relatively mild, are almost the reverse of PD's. First, she avoids nouns and overuses pronouns (spatial indexing). Second, she tends to make formational errors in root signs, producing nonsense items by substituting incorrect hand configurations, places of articulation, or movements. Thus, these two fluent patients display almost complementary deficits, breaking along linguistic fault lines, as it were, between lexicon and grammar.

The third patient (GD) is nonfluent. She has massive damage over most of the left frontal lobe, including Broca's area. She produces individual signs correctly (with her nondominant hand, due to paralysis of the right side of her body), and can repeat a test series of signs rapidly and accurately, so that her deficits are not simply motoric. Yet her spontaneous signing invites description by just those epithets that characterize a Broca's aphasic. Her utterances are slow, effortful, short, and agrammatic, largely made up of open-class items. She omits all grammatical formatives, including inflections, morphological modulations, and most spatial indices. In short, this patient, too, displays a peculiarly linguistic rather than a general cognitive pattern of breakdown.

From this brief review of brain specialization for language we may draw several conclusions. First, language breakdown seems to follow rough linguistic lines of demarcation, indicating that phonology (or patterns of sign formation) and syntax may be supported by separable neural subsystems within the left hemisphere. Second, left hemisphere specialization does not rest on a particular sensorimotor channel. Rather, the hemisphere supports general linguistic functions, common to both spoken and signed language. Thus, despite the left hemisphere's innate predisposition for speech (see below on language acquisition), its initial neural organization is sufficiently plastic to admit quite different language forms (cf. Neville, 1980; Neville, Kutas, & Schmidt, 1982). At the same time, we still do not know enough about the anatomy and physiology of the brain to be sure that areas important for particular functions in spoken language precisely correspond to areas important for analogous functions in signed language: the issue of analogy vs. homology is not yet closed.

Several further cautions should be noted. It is not yet clear (either from linguistic theory or from behavioral evidence) that syntax and phonology constitute homogeneous functions: some aspects of syntax and phonology may be separable from some aspects of language, others may not (Dennis, 1983). Second, it is even less clear that we should expect a coherent function, once specified, to be discretely and coherently localized in the brain. In looking for correspondences between one level of description (linguistic) and another level (neurological), we may be guilty of the "first-order isomorphism fallacy" that caused the downfall of phrenology and faculty psychology. The error would be analogous to that of someone who expected a single function of an automobile--say, acceleration--to be discretely and coherently localized in the engine. In fact, of course, the mechanism underlying acceleration is distributed over gears, fuel pump, carburetor, pistons, and so on. Perhaps syntactic and phonological functions emerge, like acceleration, from the coordinated actions of disparate parts.

Language Acquisition

As many as 5 percent of American children suffer from some form of delayed or disordered language development, and many more join the ranks of the illiterate. Moreover, there is growing evidence that the capacity to read depends in large part on normal development of the primary language processes of speaking and listening (Crain & Shankweiler, in press). Scientific understanding of development is therefore of broad pediatric and educational interest. In the first instance, the work may simply permit us to establish reliable norms, based on a sound understanding of what language acquisition entails. Later, we may hope, the work should lead to more effective therapeutic intervention than is now available.

No area of language study has been more strongly affected by Chomsky's work than language acquisition. Indeed, it is fair to say that until Chomsky's writings began to be widely disseminated among psychologists, in the early 1960s, the field did not exist. The few psychologists who considered the matter at all (e.g., Mowrer, 1960; Skinner, 1957) assumed that language learning would be subsumed under the general learning theory that behaviorists were striving to develop. Yet today the field has grown to such depth and complexity that a recent volume on the state of the art (Wanner & Gleitman, 1982) lists some 900 references, over half of them published in the last 10 years. The most that I can hope to do here is sketch some of the reasons for this phenomenal growth. What did Chomsky say that aroused such interest? What questions are researchers trying to answer?

Language development is a central issue in Chomsky's thought (e.g., 1965, 1972, 1980), bearing directly on the natural categories of the human mind. The issue arises from four assumptions. First, any grammar sufficient to generate the sentences of a natural language is a complex "system of many...rules of...different types organized in accordance with certain fixed principles of ordering and applicability and containing a certain fixed substructure" (1972, p. 75). Second, the descriptive predicates of this system (grammatical categories, phonological classes) are not commensurate with those of any other known system in the world or in the mind. Third, the data available to the child in the speech of others is "meager and degenerate." Fourth, no known theory of learning--least of all, a stimulus-response reinforcement theory of the kind scathingly criticized by Chomsky in his review (1959) of Skinner's <u>Verbal Behavior</u> (1957)--is adequate to account for a child's learning a language. Chomsky (1972) therefore

assigns to the mind an innate property, a schema constituting the "universal grammar" to which every language must conform. The schema is highly restrictive, so that the child's search for the grammar of the language it is learning will not be impossibly long.

Chomsky (1972) then divides the research task into three parts. First is the linguist's task: to define the essential properties of human language, the schema or universal grammar. Second is the psychologist's task of determining the minimal conditions that will trigger the child's innate linguistic mechanisms. The third task, closely related to the second, arises from the assumption that most of the utterances a child hears are not well formed. How then is the child to know which utterances to accept as evidence of the grammar it is searching for and which utterances to reject? The third task is therefore to discover the nature of the relation between a set of data and a potential grammar, sufficient to validate the grammar as a theory of the language being learned.

The proposition that language is an innate faculty of the human mind has a long history in Western thought from Plato to Darwin. The proposition is logically independent of any particular theory of language structure. Indeed, the entire enterprise of generative grammar might fail, yet leave the claim of innateness untouched. Certainly Chomsky's linguistic theories have been, and continue to be, a rich source of hypothesis and experiment in studies of language acquisition. However, his principle achievement in this area has been to force recognition that the learning of a language is an extraordinarily complex process with profound implications for the nature of mind. He has formulated the problem of language learning more precisely than ever before, spelling out its logical prerequisites in a fashion that promises to lead, given appropriate research, to a more precise specification of the innate "knowledge" that a child must bring to bear if it is ever to learn a language at all.

As we have noted, Chomsky's challenge precipitated a vast quantity of research. The first need was for data, for systematic descriptions of how language actually develops. Work initially concentrated on syntactic development (e.g., Brown, 1973), but in the past dozen years has expanded to include phonology, (e.g., Yeni-Komshian, Kavanagh, & Ferguson, 1980), semantics (e.g., Carey, 1982; MacNamara, 1982) and pragmatics (e.g., Bates & MacWhinney, 1982). As data have accumulated, it has become possible to answer many questions and, of course, to ask many more.

When does language development begin? Can we isolate reliable stages of development across children? Do the same stages occur in different language environments? Is the input to the child truly "meager and degenerate"? Is the child really constructing a grammar? Is the process passive, or must the child actively engage itself? What is the role of imitation? Do we have to posit innate proclivities? If so, are they indeed purely linguistic? And so on.

To see the force of these questions, we must have a sense of the complexity of the task that faces a child learning its native language. From our discussion of the problems of speech perception and automatic speech recognition, it will be obvious that we have much to learn about how the infant discovers invariant phonetic and lexical segments in the speech signal. We still do not know how the infant learns the basic sound pattern of a language during its first two years of life and comes to speak its first few

dozen words. But let us set these puzzles aside and go straight to early syntax, where the bulk of child language research has been concentrated. The goal of this work has been to infer from a child's utterances (performance) what it "knows" (competence) about grammar, and the meanings encoded by grammar, at each stage of its development.

Consider, as an example, the sentence cited above, I want the apple we picked for supper, a sentence comfortably within the competence of a four-year-old child. What must a child know to produce such a sentence? We will look at three aspects of its structure to illustrate the basis of Chomsky's claim that grammatical categories do not map in any simple way onto the categories of general cognition.

(1) Word order. A child who utters the sentence evidently knows the standard subject-verb-object (SVO) order of English and so says, I want the apple. The child does not say as (transposing into English) a Turkish or Japanese child might say, I the apple want (SOV) or The apple I want (OSV). Presumably, the English-speaking child has long since learned that Adam loves Eve does not mean the same as Eve loves Adam. A Turkish or Japanese child, on the other hand, would have learned that uncertainties, due to variable word order, as to the underlying relations expressed in a sentence (who does what to whom) are resolved by attaching appropriate suffixes to subject and object (Slobin, 1982).

So far, the mapping between grammar and world, in the three languages, would seem to be arbitrary but direct. However, we are given pause by another phrase in our example, the apple we picked (=the apple that we picked). Here, in an object relative clause, the order of subject (we) and object (apple) is reversed, and the verb (picked) appears at the end, giving OSV. The switch from SVO (we picked that) to OSV (that we picked) is obligatory in English object relative clauses. Notice that, to apply this rule, a child cannot draw on any knowledge of the world; rather, it must (in some sense) know the grammatical structure of the sentence. We have here, then, another example of the structure dependence, noted above in our discussion of interrogatives.

(2) Use of the article. The child says, I want the apple, not I want an apple. Of course, if many apples had been picked, an apple would have been correct. The distinction between definite and indefinite articles seems natural to an English speaker. To a speaker of Russian, Chinese, or other languages in which articles are not used, the distinction might seem tiresome and unnecessary. In fact, rules for use of articles in English are complex and, with respect to the aspects of the world that they encode, seemingly arbitrary. Yet the rules are learned by the third or fourth year of life (Brown, 1973, p. 271).

(3) Noun phrases. As a final example, consider the noun phrase, the apple we picked. These four words (article + noun + adjectival phrase) form the grammatical object of the sentence. A child who utters them must already know the general rule for constructing noun phrases in English: the adjective goes before the noun (the red apple), not, as in French, after the noun (la pomme rouge). However, there is an exception to the rule: if the adjective is itself a phrase (that is, a relative clause: (that we picked), the adjective must follow the noun (the apple we picked, not the we picked apple). Once again, the child reveals in its utterance knowledge of a rule of English grammar that cannot be derived from knowledge of the world.

In short, there are solid grounds for believing that language structure (both at the level of sound pattern, or phonology, and at the level of syntax) may be sui generis. With this in mind, let us briefly review some of what we know about the course of development, with particular attention to the questions with which we began.

The infant is biologically prepared to distinguish speech from nonspeech at, or very soon after, birth. A double dissociation of the left cerebral hemisphere for perceiving speech and of the right hemisphere for perceiving nonspeech sounds within days of birth has been demonstrated both electrophysiologically (e.g., Molfese, 1977) and behaviorally (e.g., Segalowitz & Chapman, 1980). Further, dozens of experiments in the past 10 years have shown that infants, in their first six months of life, can discriminate virtually any adult speech contrast from any language on which they are tested (e.g., [b] vs. [p], [d] vs. [g], [m] vs. [n], etc.) (Aslin, Pisoni, & Jusczyk, 1983; Eimas, 1982). There is also evidence that infants begin to recognize the function of such contrasts, to distinguish words in the surrounding language, during the second half of their first year (Werker, 1982). (For fuller review, see Studdert-Kennedy, 1986).

In terms of sound production, Oller (1980) has described a regular progression from simple phonation (0-1 months) through canonical babbling (7-10 months) to so-called variegated babbling (11-12 months). The phonetic inventory of babbled sounds is strikingly similar across many languages and even across hearing and deaf infants up to the end of the first year (Locke, 1983). These similarities argue for a universal, rather than language-specific, course of articulatory development.

However, around the end of the twelfth month, when the child produces its first words, the influence of the surrounding language becomes evident. From this point on, universals become increasingly difficult to discern, because whatever universals there may be are masked by surface diversity among languages. In this respect, the development of language differs from the development of, say, sensorimotor intelligence or mathematical ability (cf. Gelman & Brown, this volume). Nonetheless, we can already trace some regularities across children within a language and, to some lesser extent, across languages.

The most heavily studied stage of early syntactic development, in both English and some half-dozen other languages, is the so-called two-morpheme stage. Brown (1973) divides early development into five stages on the basis of mean length of utterance (MLU), measured in terms of the number of morphemes in an utterance. The stages are "not...true stages in Piaget's sense" (Brown, 1973, p. 58), but convenient, roughly equidistant points from MLU=2.00 through MLU=4.00. The measure provides an index of language development independent of a child's chronological age.

Of interest in the present context is that no purely grammatical description of Stage I (MLU=2.00, with an upper bound of 5.00) has been found satisfactory. Instead, the data are best described by a "rich interpretation," assigning a meaning or function to an utterance on the basis of the context in which it occurs. Brown lists 11 meanings for Stage I constructions, including: naming, recurrence (more cup), nonexistence (all gone egg), agent and action (Mommy go), agent and object (Daddy key), action and location (sit chair), entity and location (Baby table), possessor and possession (Daddy chair), entity and attribute (yellow block). Brown (1973)

proposes that these meanings "derive from sensorimotor intelligence, in Piaget's sense...[and] probably are universal in humankind but not...innate" (p. 201).

We should emphasize that these Stage I patterns reflect semantic, not grammatical, relations even though they may be necessary precursors to the grammatical relations that develop during Stage II (MLU=2.50, with an upper bound of 7.00). Brown (1973) traced the emergence of 14 grammatical morphemes in three Stage II English-speaking children. The morphemes included: prepositions (in, on), present progressive (I am playing), past regular (jumped), past irregular (broke), plural-s, possessive -s, third person -s (he jumps), and others. The remarkable finding was that all three children acquired the morphemes in roughly the same order (with rank order correlations between pairs of children of 0.86 or more). This result was confirmed in a study of 21 English-speaking children by de Villiers and de Villiers (1973).

However, unlike the meanings and functions of Stage I, the more or less invariant order of morpheme acquisition of Stage II has not been confirmed for languages other than English. Perhaps we should not expect that it will be. Languages differ, as we have seen, in the grammatical devices that they use to mark relations within a sentence. The devices used by one language to express a particular grammatical relation may be, in some uncertain sense, "easier" to learn than the devices used by another language for the same grammatical relation. Slobin (1982) has compared the ages at which four equivalent grammatical constructions are learned in Turkish, Italian, Serbo-Croatian, and English. In each case, the Turkish children developed more rapidly than the other children. If these results are valid and not mere sampling error, the "studies suggest that Turkish is close to an ideal language for early acquisition" (Slobin, 1982, p. 145).

Unless we suppose that Turkish parents are more attentive to their children's language than Italian, Serbo-Croatian, and English parents, we may take this result as further evidence that "selection pressures" (reinforcement) have little role to play in language learning. Brown and Hanlon (1970) showed some years ago that parents tend to correct the pronunciation and truth value, rather than the syntax, of their children's speech. Indeed, one of the puzzles of language development is why children improve at all. At each stage, the child's speech seems sufficient to satisfy its needs. Neither reinforcement nor imitation of adult speech suffices to explain the improvement. Early speech is replete with forms that the child has presumably never heard: two sheeps, we goed, mine boot. These errors reflect not imitation, but over-generalization of rules for forming plurals, past tenses, and possessive adjectives.

We come then to a guiding assumption of much current research: Learning a first language entails active search for language-specific grammatical patterns (or rules) to express universal cognitive functions. The child may be helped in this by the relative "transparency" (Slobin, 1980) of the speech addressed to it--either because the language itself, like Turkish, is transparent and/or because adult speech to the child is conspicuously well formed. Several studies (e.g., Newport et al., 1977) have shown that the speech addressed to children tends not to be "degenerate." Yet the speech may be "meager" in the sense that relatively few instances suffice to trigger recognition of a pattern (Roeper, 1982). Such rapid learning would seem to require a system specialized for discovering distinctive patterns of sound and syntax in any language to which a child is exposed.

Finally, it is worth remarking that all normal children do learn a language, just as they learn to walk. Western societies acknowledge this in their attitude to children who fail: We regard them as handicapped or defective, and we arrange clinics and therapeutic settings to help them. As Dale (1976) has remarked, we do not do the same for children who cannot learn to play the piano, do long division, or ride a bicycle. Of course, children vary in intelligence, but not until I.Q. drops below about 50 do language difficulties begin to appear (Lenneberg, 1967). Children at a given level of maturation also vary in how much they talk, what they talk about, and how many words they know. Where they vary little, it seems, is in their grasp of the basic principles of the language system--its sound structure and syntax.

Conclusion

The past 50 years have seen a vast increase in our knowledge of the biological foundations of language. Rather than attempt even a sampling of the issues raised by the research we have reviewed, let me end by emphasizing a point with which I began: the interplay between basic and applied research, and between research and theory.

The advances have come about partly through technological innovations, permitting, for example, physical analysis of the acoustic structure of speech and precise localization of brain abnormalities; partly through methodological gains in the experimental analysis of behavior; partly through growing social concern with the blind, the deaf, and otherwise language-handicapped persons. Yet these scattered elements would still be scattered had they not been brought together by a theoretical shift from description to explanation.

Perhaps the most striking aspect of the development is its unpredictability. Fifty years ago no one would have predicted that formal study of syntax would offer a theoretical framework for basic research in language acquisition, now a thriving area of modern experimental psychology, with important implications for treatment of the language-handicapped. No one would have predicted that applied research on reading machines for the blind would contribute to basic research in human phonetic capacity, lending experimental support to the formal linguistic claim of the independence of phonology and syntax. Nor, finally, would anyone have predicted that basic psycholinguistic research in American Sign Language would provide a unique approach to the understanding of brain organization for language and to testing the hypothesis, derived from linguistic theory, that language is a distinct faculty of the human mind.

Presumably, continued research in the areas we have reviewed and in related areas that we have not (such as the acquisition of reading, the motor control and coordination of articulatory action, second language learning), will consolidate our view of language as an autonomous system of nested subsystems (phonology, syntax). Beyond this lies the further task of unfolding the language system, tracing its evolutionary and ontogenetic origins in the nonlinguistic systems that surround it and from which, in the last analysis, it must derive. We would be rash to speculate on the diverse areas of research and theory that will contribute to this development.

References

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith & J. J. Campos (Eds.), Infancy and the biology of development (Vol. II, Carmichael's manual of child psychology, 4th ed.). New York: Wiley and Sons.

Baker, C., & Padden, C. A. (1978). Focusing on the nonmanual components of American Sign Language. In P. Siple (Ed.), Understanding language through sign language research (pp. 27-58). New York: Academic Press.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. R. Gleitman (Eds.), Language acquisition: The state of the art (pp. 173-218). New York: Cambridge University Press.

Bellugi, U., Poizner, H., & Klima, E. S. (1983). Brain organization for language: Clues from sign aphasia. Human Neurobiology, 2, 155-170.

Benson, D. F. (1983). Cerebral metabolism. In M. Studdert-Kennedy (Ed.), Psychobiology of language (pp. 205-211). Cambridge, MA: MIT Press.

Bever, T. G. (1970). The cognitive basis for linguistic studies. In J. R. Hayes (Ed.), Cognition and language development. New York: Wiley.

Bloomfield, L. (1933). Language. New York: Holt.

Brown, R. (1973). A first language. Cambridge, MA: Harvard University Press.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), Cognition and the development of language. New York: Wiley.

Cairns, H. S., & Cairns, C. E. (1976). Psycholinguistics. New York: Holt, Rinehart, and Winston.

Caramazza, A., & Zurif, E. B. (1978). Comprehension of complex sentences in children and aphasics: A test of the regression hypothesis. In A. Caramazza & E. B. Zurif (Eds.), Language acquisition and language breakdown (pp. 145-161). Baltimore: Johns Hopkins University Press.

Carey, S. (1982). Semantic development: The state of the art. In E. Wanner & L. Gleitman (Eds.), Language acquisition: The state of the art. New York: Cambridge University Press.

Chiba, T., & Kajiyama, M. (1941). The vowel: Its nature and structure. Tokyo: Tokyo-Kaiseikan.

Chomsky, N. (1956). Three models for the description of language. IRE Transactions on Information Theory, IT-2, 113-124.

Chomsky, N. (1957). Syntactic structures. The Hague: Mouton.

Chomsky, N. (1959). Review of Verbal behavior by B. F. Skinner. Language, 35, 26-58.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Chomsky, N. (1972). Language and mind (revised ed.). New York: Harcourt Brace Jovanovich.

Chomsky, N. (1975). Reflections on language. New York: Random House.

Chomsky, N. (1980). Rules and representations. The Behavioral and Brain Sciences, 3, 1-62.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Cole, R. A., Rudnicky, A., Reddy, R., & Zue, V. W. (1980). Speech as patterns on paper. In R. A. Cole (Ed.), Perception and production of fluent speech. Hillsdale, NJ: LEA.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. Psychological Review, 81, 348-374.

Cooper, F. S. (1950). Spectrum analysis. Journal of the Acoustical Society of America, 22, 761-762.

Cooper, F. S. (1972). How is language conveyed by speech. In J. F. Kavanagh & I. G. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading (pp. 25-45). Cambridge, MA: MIT Press.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. Journal of the Acoustical Society of America, 24, 597-606.

Cooper, F. S., Gaitenby, J. H., & Nye, P. W. (1984). Evolution of reading machines for the blind: Haskins Laboratories' research as a case history. Journal of Rehabilitation Research and Development, 21, 51-87.

Crain, S., & Shankweiler, D. (in press). Reading acquisition and language acquisition. In A. Davison, G. Green, & G. Herman (Eds.), Critical approaches to readability: Theoretical bases of linguistic complexity. Hillsdale, NJ: Erlbaum.

Dale, P. S. (1976). Language development (2nd ed.). New York: Holt, Rinehart and Winston.

Darwin, C. J. (1976). The perception of speech. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception, Vol. VII: Language and speech (p. 175). New York: Academic Press.

Dennis, M. (1983). Syntax in brain-injured children. In M. Studdert-Kennedy (Ed.), Psychobiology of language (pp. 195-202). Cambridge, MA: MIT Press.

deSaussure, F. (1966). Course in general linguistics (W. Basuin, trans.). New York: McGraw Hill.

deVilliers, J. G., & deVilliers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes. Journal of Psycholinguistic Research, 2, 267-278.

Eimas, P. D. (1982). Speech perception: A view of the initial state and perceptual mechanics. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), Perspectives on mental representation (pp. 339-360). Hillsdale, NJ: LEA.

Fant, G. (1960). Acoustic theory of speech production. 's Gravenhage: Mouton.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. Logos, 5, 3-17.

Fant, G. (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), Manual of phonetics (pp. 173-277). Amsterdam: North-Holland.

Fant, G. (1973). Descriptive analysis of the acoustic aspects of speech. In Speech sounds and features (Chapter 2). Cambridge, MA: MIT Press.

Flanagan, J. L. (1983). Speech analysis, synthesis and perception. Heidelberg: Springer-Verlag.

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: MIT Press.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). The psychology of language. New York: McGraw-Hill.

Foss, D. J., & Hakes, D. T. (1978). Psycholinguistics: An introduction to the psychology of language. Englewood Cliffs, NJ: Prentice-Hall.

Givon, T. (1979). On understanding grammar. New York: Academic Press.

Goodenough, C., Zurif, E. B., & Weintraub, S. (1977). Aphasics' attention to grammatical morphemes. Language and Speech, 20, 11-20.

Goodglass, H., & Geschwind, N. (1976). Language disturbance (aphasia). In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. 7). New York: Academic Press.

Goodglass, H., & Kaplan, E. (1972). The assessment of aphasia and related disorders. Philadelphia: Lea & Fibiger.

Hecaen, H., & Albert, M. L. (1978). Human neuropsychology. New York: Wiley.

Hockett, C. F. (1960). The origin of speech. Scientific American, 203, 89-96.

Hockett, C. F. (1968). The state of the art. The Hague: Mouton.

Jakobson, R. (1941). Kindersprache, Aphasie, und Allgemeine Lautgesetze. Stockholm: Almqvist and Wiksell.

Joos, M. (1948). Acoustic phonetics. Language, 23(S24), 1-136.

Katz, J. J. (1981). Language and other abstract objects. Totowa, NJ: Rowman and Littlefield.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 15, 166-171.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. Cortex, 8, 163-178.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In E. Bach & R. Harms (Eds.), Universals in linguistic theory (pp. 171-202). New York: Holt, Rinehart, and Winston.

Klima, E. S., & Bellugi, U. (1979). The signs of language. Cambridge, MA: Harvard University Press.

Kroeber, A. L. (1958). Sign language inquiry. International Journal of American Linguistics, 24, 1-19.

Kurath, H. (1939). Handbook of the linguistic geography of New England (with the collaboration of M. L. Hansen, J. Bloch, & B. Bloch). Providence, RI: Brown University Press.

Labov, W. (1972). Sociolinguistic patterns. Philadelphia: University of Pennsylvania.

Ladefoged, P. (1980). What are linguistic sounds made of? Language, 56, 485-502.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), Cerebral mechanisms in behavior (pp. 112-136). New York: John Wiley.

Lehmann, W. P. (1973). Historical linguistics. New York: Holt, Rinehart, and Winston.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

Lesser, R. (1978). Linguistic investigations of aphasia. New York: Elsevier.

Levi-Strauss, C. (1966). The savage mind. Chicago: University of Chicago Press.

Levinson, S. E., & Liberman, M. Y. (1981). Speech recognition by computer. Scientific American, 244, 64-76.

Liberman, A. M. (1957). Some results of research on speech perception. Journal of the Acoustical Society of America, 29, 117-123.

Liberman, A. M. (1970). The grammars of speech and language. Cognitive Psychology, 1, 301-323.

Liberman, A. M. (1982). On finding that speech is special. American Psychologist, 37, 148-167.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A. M., Ingemann, F., Lisker, L., Delattre, P. C., & Cooper, F. S. (1959). Minimal rules for synthesizing speech. Journal of the Acoustical Society of America, 31, 1490-1499.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII, Perception (pp. 143-178). New York: Springer-Verlag.

Licklider, J. C. R., & Miller, G. (1951). The perception of speech. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley.

Liddell, S. K. (1978). Nonmanual signals and relative clauses in American Sign Language. In P. Siple (Ed.), Understanding language through sign language research (pp. 59-90). New York: Academic Press.

Lieberman, P., & Crelin, E. S. (1971). On the speech of Neanderthal man. Linguistic Inquiry, II, 203-222.

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. American Anthropologist, 74, 287-307.

Limber, J. (1973). The genesis of complex sentences. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press.

Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. Cognition, 13, 361-392.

Locke, J. L. (1983). Phonological acquisition and change. New York: Academic Press.

Luria, A. R. (1966). Higher cortical functions in man. New York: Basic Books.

Luria, A. R. (1970). Traumatic aphasia. The Hague: Mouton.

MacNamara, J. (1982). Names for things. Cambridge, MA: MIT Press.

Mattingly, I. G. (1968). Experimental methods for speech synthesis by rule. IEEE Transactions on Audio and Electroacoustics, AU-16, 198-202.

Mattingly, I. G. (1974). Speech synthesis for phonetic and phonological models. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 12, pp. 2451-2487). The Hague: Mouton.

Mayberry, R. I. (1978). Manual communication. In H. Davis & S. R. Silverman (Eds.), Hearing and deafness (4th ed.). New York: Holt, Rhinehart, & Winston.

Mayr, E. (1974). Behavior programs and evolutionary strategies. American Scientist, 62, 650-659.

Miller, G. A. (1951). Language and communication. New York: McGraw-Hill.

Miller, G. A., Galanter, E., & Pribam, K. H. (1960). Plans and the structure of behavior. New York: Henry Holt and Co., Inc.

Molfese, D. L. (1977). Infant cerebral asymmetry. In S. J. Segalowitz & F. A. Gruber (Eds.), Language development and neurological theory. New York: Academic Press.

Moscovitch, M. (1983). Stages of processing and hemispheric differences in language in the normal subject. In M. Studdert-Kennedy (Ed.), Psychobiology of language (pp. 88-104). Cambridge, MA: MIT Press.

Mowrer, O. H. (1960). Learning theory and the symbolic processes. New York: John Wiley and Sons.

Müller, J. (1848). The physiology of the senses, voice, and muscular motion with the mental faculties (W. Baly, Trans.). New York: Walton and Maberly.

Neville, H. J. (1980). Event-related potentials in neuropsychological studies of language. Brain and Language, 11, 300-318.

Neville, H. J., Kutas, M., & Schmidt, A. (1982). Event-related potential studies of cerebral specialization during reading. II. Studies of congenitally deaf adults. Brain and Language, 16, 316-337.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Snow & C. Ferguson (Eds.), Talking to children: Language input and

acquisition. Cambridge: Cambridge University Press.

Oettinger, A. (1972). The semantic wall. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw-Hill.

Ojemann, G. (1983). Brain organization for language from the perspective of electrical stimulation mapping. The Behavioral and Brain Sciences, 6, 218-219.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology Vol. 1: Production (pp. 93-112). New York: Academic Press.

Pick, A. (1913). Die agrammatischen sprachstorungen. Berlin: Springer.

Porter, R. J., Jr., & Hughes, L. F. (1983). Dichotic listening to CVs: Method, interpretation and application. In J. Hellige (Ed.), Cerebral hemispheric asymmetry: Method, theory, and application. New York: Praeger.

Potter, R. K., Kopp, G. A., & Green, H. G. (1947). Visible speech. New York: Van Nostrand.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. The Behavioral and Brain Sciences, 3, 111-169.

Reddy, D. R. (1975). Speech recognition: Invited papers at the 1974 IEEE Symposium. New York: Academic Press.

Roeper, T. (1982). The role of universals in the acquisition of gerunds. In E. Wanner & L. R. Gleitman (Eds.), Language acquisition: The state of the art. New York: Cambridge University Press.

Seashore, R. H., & Erickson, L. D. (1940). The measurement of individual differences in general English vocabularies. Journal of Educational Psychology, 31, 14-38.

Segalowitz, S. J., & Chapman, J. S. (1980). Cerebral asymmetry for speech in neonates: A behavioral measure. Brain and Language, 9, 281-288.

Shankweiler, D., & Studdert-Kennedy, M. (1967). Identification of consonants and vowels presented to the left and right ears. Quarterly Journal of Experimental Psychology, 19, 59-63.

Skinner, B. F. (1957). Verbal behavior. New York: Appleton-Century Crofts.

Slobin, D. I. (1980). The repeated path between transparency and opacity in language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie.

Slobin, D. I. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. Gleitman (Eds.), Language acquisition: State of the art. New York: Cambridge University Press.

Stevens, K. N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech (pp. 303-330). New York: Academic Press.

Stevens, K. N., & House, A. S. (1955). Development of a quantitative description of vowel articulation. Journal of the Acoustical Society of America, 27, 484-493.

Stevens, K. N., & House, A. S. (1961). An acoustical theory of vowel production and some of its limitations. Journal of Speech and Hearing Research, 4, 303-320.

Stokoe, W. C., Jr. (1960). Sign language structure. In Studies in Linguistics: Occasional Papers 8. Buffalo: Buffalo University Press.

Stokoe, W. C., Jr. (1974). Classification and description of sign languages. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. XII, pp. 345-371). The Hague: Mouton.

Stokoe, W. C., Jr., Casterline, D., & Croneberg, C. (1965). A dictionary of American Sign Language. Washington, DC: Gallaudet College Press.

Studdert-Kennedy, M. (1974). The perception of speech. In T. A. Sebeok (Ed.), Current trends in linguistics (Vol. 9). The Hague: Mouton.

Studdert-Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.), Contemporary issues in experimental phonetics (pp. 243-293). New York: Academic Press.

Studdert-Kennedy, M. (Ed.) (1983). Psychobiology of language. Cambridge, MA: MIT Press.

Studdert-Kennedy, M. (1986). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: LEA.

Studdert-Kennedy, M., & Lane, H. (1980). Clues from the differences between signed and spoken language. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form (pp. 29-39). Deerfield Park, FL: Verlag Chemie.

Studdert-Kennedy, M., & Shankweiler, D. (1970). Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 48, 579-594.

Templin, M. C. (1957). Certain language skills of children. Minneapolis: University of Minnesota Press.

Von Stockert, T. (1972). Recognition of syntactic structure in aphasic patients. Cortex, 8, 323-335.

Wanner, E., & Gleitman, L. R. (Eds.) (1982). Language acquisition: The state of the art. New York: Cambridge University Press.

Werker, J. F. (1982). The development of cross-language speech perception: The effect of age, experience and context on perceptual organization. Unpublished doctoral dissertation, University of British Columbia.

Wilson, E. O. (1975). Sociobiology. Cambridge: The Belknap Press.

Yeni-Komshian, G. H., Kavanagh, J. F., & Ferguson, C. A. (Eds.) (1980). Child phonology (Vols. 1 and 2). New York: Academic Press.

Zaidel, E. (1978). Lexical organization in the right hemisphere. In P. A. Buser & A. Rougeul-Buser (Eds.), Cerebral correlates of conscious experience (pp. 177-197). Amsterdam: Elsevier/North Holland.

Zaidel, E. (1980). Clues from hemispheric specialization. In U. Bellugi & M. Studdert-Kennedy (Eds.), Signed and spoken language: Biological constraints on linguistic form. Weinheim: Verlag Chemie.

Zaidel, E. (1983). On multiple representations of the lexicon in the brain--the case of two hemispheres. In M. Studdert-Kennedy (Ed.), Psychobiology of language (pp. 105-125). Cambridge, MA: MIT Press.

Zurif, E. B., & Blumstein, S. E. (1978). Language and the brain. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), Linguistic theory and psychological reality (pp. 229-245). Cambridge, MA: MIT Press.

THE PURSUIT OF INVARIANCE IN SPEECH SIGNALS*

Leigh Lisker†

Abstract. The search for the acoustic properties useful to the listener in extracting the linguistic message from a speech signal is often construed as the task of matching invariant physical properties to invariant phonological percepts; the discovery of the former will explain the latter. These phonological percepts are essentially the phonemes of pregenerative phonology, and they are more or less faithfully reflected in standard alphabetic writing. Thus English deep and doom are supposed to be perceptually identical in their initial /d/s; the orthographic similarity is in agreement with the linguist's "representation" of these forms. The partial identity in spelling is only weak evidence for perceptual invariance, however. First, while some phonemes may comprise a single "sound," others are said by linguists to include phonetically distinct ones. Thus English /p/ includes both aspirated and unaspirated voiceless labial stops. The view that it is not the phoneme, but rather the phonetic feature, to which an acoustic invariant might be attributed, raises two questions: (a) Since segments sharing a feature are rarely judged to constitute a single sound, the search for a feature-specific invariant, whose function is to explain perceptual constancy, is deprived of its essential motivation, and (2) there is no more reason to expect the acoustic cues to a feature to be context-independent than is the case with the phoneme. What seems more likely is to find that some phonemes, and some features, are more invariantly marked in the speech signal than others.

The auditory analysis of speech into sequences of elementary speech sounds long antedates the development of our present methods for the instrumental recording and analysis of acoustic signals. The alphabetic registration of speech, and, in particular, its phonetic and phonological spellings by linguists, embody a once generally accepted model for signals produced and perceived in the speech communication process: Speech is articulated, that is, jointed, so that a sequence of discrete vocal tract shapes gives rise to a sequence of similarly discrete sounds, which, in turn, is interpreted as some specific linguistic message. In some part, this view still prevails. Speech is now regarded as being both articulated and fluent, and we continue to look for acoustic properties by which each category of phonetic segments, or the phonological unit to which it is assigned, may be

---

*Journal of the Acoustical Society of America, 1985, 77, 1199-1202.
†Also Department of Linguistics, University of Pennsylvania.
Acknowledgment. Preparation of this paper was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]

characterized. We persist, moreover, in thinking of these sought-after properties as attributes of discrete and acoustically delimitable intervals to which the names of our phonetic/phonological categories are directly applicable, thereby conflating the rather different units designated by the terms "phonetic segment" (or "speech sound") and "acoustic segment" (see e.g., Repp, 1981).

Surveys of the modern literature addressing the invariance question (e.g., Cooper, 1980; Darwin, 1976; Liberman & Studdert-Kennedy, 1978; Wickelgren, 1976) suggest that neither the definition of invariance nor the type of linguistic unit to be specified by physical invariants has held constant. Invariance has been posited, sometimes to be dismissed, but sometimes perhaps demonstrated with convincing plausibility, at several levels of abstraction--as a temporal interval having a "typical" waveform (Fletcher, 1929), a particular spectral property (Stevens & Blumstein, 1978) or a given dynamic pattern (Kewley-Port, 1983), by a set of "target" formant frequencies (Lindblom & Studdert-Kennedy, 1967), or by so-called "locus" frequencies (Delattre et al., 1964). Moreover, there does not seem to be entire agreement as to either the size or level of abstractness of the linguistic elements for which invariant acoustic properties (given some definition of "invariant") are to be sought; Should they be phonetic features, segments, demisyllables, or syllables? For any one of these entities, at what level of abstractness should they be construed? Clearly, unless there is agreement on these matters, we cannot pose the problem of invariance so that it can be resolved. Even with such agreement it is by no means self-evident that a single answer will ever be forthcoming, one that is valid for all elements of the same size and level of abstractness.

In considering the invariance question, we must remember that the original motivation of the search for acoustic invariants was to explain why speech signals can be perceived as sequences of "sounds" drawn from a limited inventory of such elements, whose freedom to occur in a virtually unlimited number of combinations makes human speech and language possible. The perceptual invariance that presumably characterizes each sound type is of a special kind--it is not auditory invariance, but only invariance with respect to those auditory properties that have what we might call potential linguistic significance, or perhaps phonetic significance. In short, the members of a sound type share the property of phonetic invariance, and one way of construing the invariance problem is to specify it as a task of determining what acoustic invariants, if any, can be associated with each of the elements for which phonetic invariance is posited. In recent years, however, emphasis has been shifted from the segment to the phonetic feature as the linguistic element to be paired with an acoustic invariant. This shift, although it faithfully reflects the practice of current phonological analysis, has at least one serious drawback--namely, that, even if a feature can be associated with an acoustically invariant property, the feature is a component of a phonetic segment (which is not abolished), and segments sharing this feature do not constitute a perceptually invariant set unless they are identical in respect to all their constituent features. But the "bundle" of all these features is the segment. Thus the smallest size unit for which (phonetic) perceptual invariance can be claimed is not the feature, but the segment, and the most abstract category level of this size and perceptual status is the phoneme of pregenerative phonology.

In the discussion of a possibly invariant relation between phonetic and acoustic properties, we must bear in mind that the first question for the linguist is not one of evaluating the similarity relations among segments, but of deciding, with respect to the speech events observed in a language community, which of them, taken pairwise, are perceived by community members to be repetitions of each other, and which are not. If their behavior leads the linguist to suppose that two events are functionally the same, then the linguist may decide that they are phonologically identical, that is, composed of the same segments in the same order. But if two events are judged to be functionally and perceptually different for the language community, then the linguist cannot on the same basis decide whether they are in part the same for speakers of the language. Because there can be no experimental verification of the perceptual identity or nonidentity of two phonetic segments in different contexts that is nearly as direct as can be applied in deciding the relation between speech events, the establishment of a collection of segments abstracted from different events as a phonetic or phonological category rests on auditory and linguistic judgments by the linguist, judgments that include hypotheses about the native speaker's perceptions of the segments. Thus the linguist can readily decide by test that the English forms deep and doom are phonetically distinct, but not whether, for the native speaker, they are identical in their initial consonants and different in their vowels and final consonants.

It might be supposed that the similarity in the linguist's spellings of deep and doom reflects a perceptual invariant for which an acoustic invariant awaits discovery. A partial identity in spelling, however, is a doubtful basis for anticipating acoustic invariance, for we might suppose the asserted identity of the two words to be as much dependent on the difference in their contexts (on the analogy of a modified Mueller-Lyer Illusion) as on the presence of a common acoustic property. The words calf and cough are also alike in the phonological spelling of their initial consonants and different in their vowels, i.e., /kæf/ and /kɔf/. A speaker of Arabic, however, might dispute this way of representing the nature of the contrast, equating calf with Arabic كاف and cough with قاف, and claiming that the difference resides ("contrastively") in the initial consonants and not in the vowels. The observing linguist, equally conversant in or perhaps equally ignorant of both languages, would say that, in the two word pairs, the phonetic differences involve both the consonants and the vowels. Thus the speech researcher, in quest of acoustic invariants matching the phonological units represented in spelling, whether standard orthographic or phonemic, could define the task variously, depending on whether he or she wanted to account acoustically for the phonologically defensible spelling behavior of the English speaker, the Arabic speaker, or the linguist. The latter would not only be of the opinion that the words in both languages differ in the initial consonants and in the vowels, but that English cough and Arabic قاف are far from being the same in their initial consonants. From all this, then, we are entitled to believe that the degree of invariance by which the onsets of deep and doom are connected is not the same as that linking the two initial consonants of calf and cough. (We may recall from these examples the findings of Liberman et al., 1952, and Schatz, 1954, that indicate that English /d,t/ are more nearly invariant in their burst than either /b,p/ or /g,k/.)

Additional examples from English can be cited that do not encourage us to expect to find invariant acoustic properties marking the phonological categories commonly recognized. The ability of listeners to distinguish the

words beeper and peeper is ascribed entirely to the /b/-/p/ contrast, /b/ being characterized usually as [+ voice] and /p/ as [-voice]. The medial /p/ of both words is, of course [-voice]. But, while it is no doubt correct to say that initial /b/ is more voiced than initial /p/, it is not so clear that it is regularly more voiced than medial /p/. Thus in a phrase this beeper the two labial stop consonants need differ not at all in degree of voicing, certainly never as much as do the stops in this peeper. Moreover, a pair of expressions, this beaker and the speaker, if they are said to include a /b/ and a /p/, respectively, can certainly not be distinctively marked by invariant acoustic properties associated with the stop voicing contrast.

The notorious writer-rider pair of many varieties of American English is another case that poses a problem. If the phonemes /t/ and /d/ are to be associated with invariants marking, respectively, the word sets tear toll heat rote and dear dole heed road, then the inclusion of writer in the first set and rider in the second must be at the cost of any claim that /t/ and /d/ are distinctively and invariantly marked. (Since some British English speakers use a voiceless aspirated stop in writer, we must accept as fact that in American English the /t/-/d/ contrast, if it operates to separate writer and rider, is marked in a less than maximally invariant fashion.) When I asked linguistically untrained speakers their opinion as to the basis on which they distinguished the two words, I failed to elicit answers consistent enough to justify a conclusion that (1) the first vowels are different perceptually and the medial consonants are identical, or (2) the vowels are the same and the consonants distinct, or (3) both vowels and following consonants are perceived as different. Under this kind of questioning, moreover, those listeners who first opted strongly for some one view soon enough showed all the uncertainty that experienced linguists have expressed over the many years that this troublesome pair of words has been a subject of dispute (see, e.g., Fischer-Jørgensen, 1975; Hymes & Fought, 1975).

The writer-rider example might be faulted as irrelevant to the present discussion precisely on the ground that listeners do not agree on what they hear as different when they distinguish auditorily between the two words. Absent such agreement, we may continue to posit an acoustic basis for connecting writer with write and rider with ride, but we need not assume that the identification of the flap in writer with /t/ and the one in rider with /d/ is based on segment-specific invariant properties. The phonemic encodings of writer rider as, e.g., /raytər/ /raydər/ are dictated by considerations that include no strong claim about the perceptual status of the alveolar flaps in those words. Hence, the motivation for seeking invariant properties connecting them "correctly" with /t/ and /d/ is weak, if not entirely lacking.

Another case involving the voicing contrast does have more relevance to the invariance question; this is the case of the post-/s/ stops in word-initial position in English. If we believe that the linguist's spelling of spin is evidence that the stop is perceived as a member of /p/, then we might describe the effect of replacing the /s/-noise with silence as one of shifting /p/ to /b/ (see Lotz et al., 1960). On the other hand, replacing the closure voicing in a token of the word ruby with silence of a certain (i.e., greater) duration will often cause listeners to report having rupee instead (Lisker, 1957a). Thus silence in one context is a "cue" to /b/, in another to /p/. There are, one would agree, other ways of describing this situation, but none will entirely explain away the problem it poses for a claim that the /p/-/b/ contrast is correlated with an acoustically invariant difference.

It may be appropriate to recall that the phonological literature was once alive with controversy as to whether the English stops are distinctively voiced and voiceless, with aspiration a redundant feature of some members of the voiceless category, or whether, instead, they are distinctively weak (lenis) and strong (fortis) in force of articulation, with voicing a redundant feature of the weakly articulated category (see e.g., Jakobson & Waugh, 1979). If the voicing of /b,d,g/ is disposable in initial and some other positions, and if aspiration is positively unnatural except initially and preceding the stressed vowel of a word, then we may claim that the /b,d,g/-/p,t,k/ contrast is signaled only by "redundant" features. If such a claim is dismissed as simply too "radical" to be considered seriously, the claim that membership in the /b,d,g/ and /p,t,k/ sets is definable in terms of acoustic invariants seems to revive a notion that is widely thought to have been conclusively demolished by the generative phonologist--namely, the biuniqueness relation between phonetic segment and phonological category (Chomsky & Halle, 1968).

The case of stop voicing involves the relation between acoustic and linguistic/perceptual aspects of the speech signal. A similar relation between articulation and linguistic percept can also be suggested. The two events represented as /iwi/ and /uyu/ in English involve the glides /w/ and /y/, the first described as tongue backed and lip rounded, the second as tongue fronted and lip unrounded. It is possible, however, to produce a recognizable /iwi/ without moving the tongue from an /i/ position, and to produce an /uyu/ without moving the lips from a posture appropriate to /u/. The vocal-tract shapes to and from which the glides are articulated are the same for these perhaps unusual ways of producing /iwi/ and /uyu/; that configuration is the one used in pronouncing the French front rounded glide of the word huit [ɥit]. I confess that I have not been able to produce these sequences so that the two lowest formants show exactly the same frequencies at the midpoints of the glides, and my claim as to the articulations should be checked by x-ray monitoring. However, my claim is no more doubtful, I would submit, than many another description of articulation for which no evidence other than proprioceptive introspection by the linguist speaker is provided. There are, moreover, "harder" data from experiments in synthesis to show that the same set of formant frequencies in different vowel-like contexts will be reported as more than one member of the /w,r,l,y/ set, e.g., as iri ala uyu (Lisker, 1957b).

In conclusion, it can be said that the search for acoustic properties by which linguistic messages are signaled in speech should and will continue to be vigorously pursued, for this enterprise is, after all, a central one in phonetics. To the extent that invariant correlates of those linguistic units having the status of perceptually defined elements turn up, fine. In some cases these elements may well be the phonemes of pregenerative phonology. But these phonemes, which linguists and the rest of us recognize in our various spelling practices, are not all perceptual constants, and we must therefore be prepared to find that some phonemes are less invariantly marked than others. If the site of acoustic invariance is postulated to be the phonetic feature rather than the phoneme, then we must still reckon with the likelihood that some features, e.g., voicing, are acoustically less stable across contexts than others, e.g., nasality. In other words, we should be prepared to live with the finding that acoustic invariance is itself a variable.

References

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper & Row.

Cooper, F. S. (1980). Acoustics in human communication: Evolving ideas about the nature of speech. Journl of the Acoustical Society of America, 68, 18-21.

Darwin, C. J. (1976). The perception of speech. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception: Vol. VII. Language and speech (p. 175). New York: Academic Press.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1962). Formant transitions and loci as acoustic correlates place of of articulation in American fricatives. Studa Linguistica, 16, 104-121.

Fischer-Jórgensen, E. (1975). Trends in phonological theory. Copenhagen: Akademisk Forlag.

Fletcher, H. (1929). Speech and hearing. New York: Van Nostrand.

Hymes, D., & Fought, J. (1975). American structuralism. In T. A. Sebeok (Ed.), Current trends in linguisics, Vol. 13: Historiography of linguistics. Mouton: The Hague.

Jakobson, R., & Waugh, L. (1979). The sound shape of language. Bloomington, IN: Indiana University Press.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. Journal of the Acostical Society of America, 73, 322-335.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus variables. American Journal of Psychology, 65, 497-516.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. Leibowitz, & H. L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII, Perception (pp. 143-178). New York: Springer-Verlag.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant-transitions in vowel recognition. Journal of the Acoustical Society of America, 42, 830-843.

Lisker, L. (1957a). Closure duration and the intervocalic voiced-voiceless distinction in English. Language, 33, 42-49.

Lisker, L. (1957b). Minimal cues for separating /w,r,l,y/ in intervocalic position. Word, 13, 257-267.

Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., & Nemser, W. (1960). The perception of English stops by speakers of English, Spanish, Hungarian, and Thai: A tape-cutting experiment. Language and Speech, 3, 71-77.

Repp, B. H. (1981). On levels of description in speech research. Journal of the Acoustical Society of America, 69, 1462-1464.

Schatz, C. (1954). The role of context in the perception of stops. Language, 30, 47-56.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

Wickelgren, W. A. (1976). Phonetic coding and serial order. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception: Vol. VIII. Languae and spech. New York: Academic Press.

HOW IS THE ASPIRATION OF ENGLISH /p,t,k/ "PREDICTABLE"?*

Leigh Lisker†

Abstract. Aspiration as a phonetic property of the English stop categories is usually said to be nondistinctive on the ground that its occurrence can be accounted for by context-sensitive rules. The word-pair pin-spin is often cited by way of example. The word-initial voiceless stop is aspirated; the post-/s/ voiceless stop is not. But the presence of aspiration is "predicted" only for some voiceless stops--namely those that are "spelled" phonologically /p/ and are either word-initial or in a position where the next vowel is stressed and in the same word. Initial stops that are spelled /b/, as in bin, may also be voiceless, so that a rule that predicts aspiration from the voicelessness of an initial stop will not work, since bin is never aspirated. Thus the knowledge on which the prediction is based is not the voicelessness of the stop, or indeed on any other ascertainable phonetic property. We know that in some words voiceless initial stops can be freely replaced by voiced stops without semantic effect, and that those voiceless stops are never aspirated, while in other words there are initial voiceless stops that are regularly aspirated, and cannot be freely replaced by voiced stops. In other words, we know whether a voiceless stop is to be aspirated or not if we know how it is spelled phonologically.

Few if any introductory linguistics textbooks in English address the subject of phonology without referring to the two kinds of p said to occur in words such as pin and spin, the first characterized by a feature of aspiration absent from the second. In a phonetic spelling of the forms, the two are commonly represented as [p^h] and [p]. Whether the phoneme /p/ is produced with or without aspiration is said to be determined by context, or, in current parlance, to be predictable by rule, this feature being present when /p/ is word-initial, but absent if a word-initial /s/ precedes it. The aspiration is then termed redundant, and moreover, so the argument often goes, it never serves as the sole basis by which lexical distinctions are signaled in English (thus Akmajian, Demers, & Harnish, 1979; Anderson, 1974; Fromkin & Rodman, 1983). Phonologists seem not to have very clearly decided whether or not this redundant feature makes some (or even a major, cf. Hyman, 1975) contribution to the auditory identification of the speech signal, nor might they all agree that the point should be decided on the basis of empirical data. These matters, while deserving discussion, are not at issue in this letter.

---

*Letter to the Editor, Language and Speech, 1985, 27, 391-394.
†Also University of Pennsylvania.
Acknowledgment. This work was supported in part by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]

The view that the aspiration observed in pin ([p^hɪn]) is irrelevant to the phonological representation of the word appears to depend on the acceptability of certain other assertions about pin and spin. First of all, it would seem that we must unquestioningly accept the labial stop of spin as a member of the /p/ phoneme, despite the recognized fact that in the position following a word-initial /s/ the so-called "p" has no distinctive status as a member of the /p/ rather than the /b/ phoneme; either a form /sbɪn/ or a form /spɪn/ is possible in English, but while there is for most phonologists a theoretical motivation for choosing at least one of them, there exists none for preferring one over the other, or for positing both. The status of the stop in spin as /p/ seems to rest on little more than the spelling convention of standard orthography, one that is simply copied in the linguist's representation. To appeal to the phonetic difference(s) between the stops of pin and spin as the basis for the redundancy of aspiration is to construct a rather flimsy argument, one that any reasonably alert beginning student might be expected to question. However, though the argument is a poor one, a more convincing case for the redundant status of aspiration is easily made, since the sound type [p] also occurs in contexts where it is distinct from [b], e.g., in rapid (vs. rabid). Moreover, a comparison of rapid with rapidity gives additional motivation for assigning [p] and [p^h] to the same phoneme, and thus for discounting the phonological significance of aspiration. In any event /p/ may be said to have both aspirated and unaspirated varieties, though to base this conclusion on the relation between pin and spin is pedagogically unfortunate.

The "predictability" of aspiration as a feature of word-initial /p/ is said to rest on the fact that /p/ is [-voiced] (e.g., Schane, 1973). Since, in point of fact, word-initial /b/ is often no more voiced than the labial stops of spin or rapid, it must be acknowledged that it is simply false to say that word-initial voiceless stops are regularly followed by aspiration. If phonologists did not persistently transcribe bin as [bɪn] and [b̥ɪn], but instead more straightforwardly wrote [bɪn] and [pɪn], the matter would be quite obvious. (Some observers have claimed that initial /b/ is not voiceless, but only "devoiced" or "partially voiced," e.g., Trager & Smith. 1951, Ladefoged, 1982, but this seems more an effort to justify writing it [b̥] for phonological reasons than to capture any phonetic difference between this /b/ and the stop in spin or rapid.) It would, however, lead students, in comparing bin = [pɪn] with pin = [p^hɪn] (or [phɪn]), to wonder about the redundant nature of the aspiration. What is true about the relation between voicing and aspiration is that a word-initial voiced stop is never followed by aspiration in English. Therefore, we can say that the presence of aspiration following a word-initial stop release allows us to infer the absence of pre-release voicing, though the absence of aspiration is compatible with both [+voiced] and [-voiced] closure. Thus, insofar as the presence or absence of one phonetic feature of the stop is to be predicted on the basis of another, we can state the rules as

[+aspirated] → [-voiced](=/p/)

and equivalently, by modus tollens

[+voiced] → [-aspirated](=/b/)

The phonological status of a stop that is [-voiced] and [-aspirated] is undecidable except on paradigmatic grounds, that is, on the basis of its contrasting with another homorganic stop. The [p] of bin is /b/ because it contrasts with the [p^h] of pin, while the [p] of rapid is /p/ by virtue of the phonologically unambiguous [b] of the contrasting rabid. The [-voiced] stop in the first word is not subject to the aspiration rule because it is assigned to the phoneme /b/, while the one in the second is not because its context makes the rule inapplicable. The stop of spin is not only [-voiced, -aspirated], and therefore of ambiguous phonological affiliation on phonetic grounds, but its status as between /p/ and /b/ cannot be decided on the basis of its contrasting with any stop that is either [+aspirated] (therefore /p/) or [+voiced] (and therefore /b/).

Of course these rules presuppose knowledge of two other kinds of information: 1) the location of word boundaries, which are not in general signaled phonetically, and 2) the location of "phonetic" segment boundaries, which are also determined by phonological considerations. In the absence of the first kind of information, no statement that either aspiration or voicing is phonologically redundant has validity, since (because there is the phoneme /h/) each feature freely occurs both with and without the other, with no third feature (i.e., stress) as a constraining factor. In the absence of phonological knowledge, on the basis of which */b^h/ and */d^h/ are not included in the English phoneme inventory, we should either have to exclude forms such as abhor and adhere from the English lexicon or consider the rule given above to be invalid. (A complicating fact is that the aspiration itself takes two forms, a voiceless one after a voiceless interval, and a voiced or murmured one after a voice interval. The latter variety is never evaluated as a stop feature in English.)

The conclusion to be drawn from the points just presented is that the predictability of the aspiration feature of the English stops is not phonetically based. Neither its presence nor its absence hinges entirely on the presence or absence of any other phonetic feature. If we know that a stop is voiceless and does not form a cluster with a preceding /s/, and if we know that it is word-initial or that the next vowel is stressed and within the same word, and if we know that it is spelled phonologically /p/ and not /b/, then we can infer that its release will be aspirated. The absence of aspiration can be predicted, given a voiceless closure, from the knowledge that it is written phonologically as /b/, or that, if /p/, a following vowel is either unstressed and in the same word or is separated from the stop by a word boundary. Finally, the rule according to which /p/ is [-aspirated] after a word-initial /s/ is no more "interesting" than another possible rule, one of broader applicability, according to which /b,d,g/ are generally [-voiced] following any voiceless obstruent, without regard to word boundary. In other words, on phonetic grounds the so-called /p,t,k/ in post-/s/ position might just as plausibly be derived by a devoicing rule applied to underlying /b,d,g/ as by a deaspirating rule applied to /p,t,k/, that is, provided the phonologist is willing to define the underlying /b,d,g/ as [+voiced, -aspirated] and the underlying /p,t,k/ as [-voiced, +aspirated]. The native speaker knows when to aspirate an initial voiceless stop and when not to, but the stop is not aspirated because it is voiceless and initial; rather it is voiceless because it is aspirated. To produce an intelligible and "normal" pin, the native speaker knows (s)he must aspirate the stop, and this precludes any voicing; for bin (s)he knows aspiration would be a mistake, but voicing is ad libitum.

References

Akmajian, A., Demers, R. A., & Harnish, R. M. (1979). Linguistics: An introduction to language and communication. Cambridge, MA: MIT Press.

Anderson, S. R. (1974). The organization of phonology. New York: Academic Press.

Fromkin, V., & Rodman, R. (1983). An introduction to language (3rd ed.). New York: Holt, Rinehart, and Winston.

Hyman, L. M. (1975). Phonology: Theory and analysis. New York: Holt, Rinehart, and Winston.

Ladefoged, P. (1982). A course in phonetics (2nd ed.). New York: Harcourt Brace.

Schane, S. A. (1973). Generative phonology. Englewood Cliffs, NJ: Prentice-Hall

Trager, G., & Smith, H., Jr. (1951). An outline of English structure (Studies in Linguistics: Occasional Papers 3). Norman, OK: Battenburg Press.

DEVELOPMENTAL PHONOLOGY: IS THE CHILD FATHER TO THE MAN?*

Catherine T. Best†

Locke's basic premise for this monograph is his "...belief that language acquisition can be understood--not merely described--and that...phonological development and change are dynamic processes in which cognitive, biological, and social factors continuously interact throughout the life of human speakers (p. xiii)." That prefatory statement is quite apropos of the book. It reflects not only the substance but also the form of the discussion, revealing both strengths and certain weaknesses. As it suggests, the psycholinguistic contribution of the work lies in the vast evidence marshalled toward the central goal of delineating the forces behind phonological growth. Of interest to developmental psychologists are its perspective that developmental processes continue throughout the lifespan, and that phonological ontogeny is shaped by the interaction of biological (intrinsic) and environmental (extrinsic) forces. But the prefatory statement also foreshadows recurrent problems in the book. First, it implies that other students of language acquisition take a merely descriptive approach, which would come as some surprise to established writers on this topic such as Bloom, Greenfield, Ferguson, Menn, Nelson, and many others. Thus, we get the semblance of a straw man, and no sense that others besides Locke believe language acquisition can be understood. Second, the book's interactionist perspective sounds grand in the abstract but falls short of adequate explanatory power, since it remains too abstract and arrives ex post facto. I will discuss these points further after a brief summary of the book's organization and contents.

Overview

At its core, the book is an extensive, annotated review of phonological and phonetic studies on various groups of people under a variety of conditions. This literature is used to discern parallel phonological characteristics between child and adult speech, which serve as the grist for two arguments about direction of causal influence: first, that intrinsic tendencies in the infant and child form the basis for adult phonological patterns and change (chapters 1-4); second, that influences are also visited upon the child from adult phonological behavior (chapters 5-6). Chapter 1 asks the question "When does phonology begin?" and answers "Before the first words," based on the restricted range and skewed distribution of phonemic elements transcribed from infant babbling. The universality of this pattern is taken as evidence of an underlying physiological basis for infants'

---

*Review of Locke, J. L. (1983). Phonological acquisition and change. New York: Academic Press.

†Also Wesleyan University.

Acknowledgment. This work was supported in part by NICHD Grant No. HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]

phonetic tendencies. Chapter 2 poses the related question, "When does phonological acquisition begin?" Its cross-language review of phonological research on early language acquisition reveals that the universal tendencies continue to shape the child's early words. These tendencies are not bent toward the phonological particulars of the native language until the final stage in a proposed three-stage model of phonological development, the "systemic" stage that presumably begins when the child has acquired a roughly 50-word vocabulary. Chapter 3 finds the intrinsic phonetic tendencies alive and well in a wide array of adult speech contexts--casual conversation, lexical avoidance, slips of the tongue, inebriation, neurological dysfunctions, glossolalia, historical sound change, and phonological universals. As summarized in Chapter 4, they are evident, as well, in the phonetics, phonotactics, and phonemic distributions within the lexicons of modern languages. Since "[t]he language and the child must both be in the equation, as each is under scrutiny (p. 186)," Chapter 5 asks "What is the child's actual phonological environment?" It considers the potential effects of adult phonetic variability upon the child's phonological development, including the extreme case of language death. The sixth and final chapter discusses the interaction between child and language by reconsideration of phonological changes (phonologization, dephonologization, rephonologization) within individual ontogeny and within the evolution of particular languages.

Evaluation

The monograph is quite commendable in a number of respects. First and foremost, it is a remarkably broad-ranging compendium of findings, which presents more comprehensively than elsewhere the universal phonological properties and phonetic tendencies observed in children and adults. It raises a variety of thought-provoking questions, and points out several intriguing between-group parallels in speech behavior, such as that between infant phonetic proclivities and the phonotactic constraints and distributions of phonemic elements found in glossolalia. As a developmental psychologist, I was attracted to the view of children as active contributors to phonological processes within a language, as opposed to their more traditional treatment as passive acquisitors or recipients of some immutable adult language. Also appealing was the argument that actual adult speech must serve as the linguistic model for children, rather than the usual assumption that their source of reference is the linguist's ideal representation of the language. In addition, as a biopsychologist I particularly appreciated the attempt to trace the observed phonetic tendencies to a biological substrate, and the evidence of continuity from prelinguistic infancy into later periods of language use.

There are, however, some notable drawbacks to the book. For one, it seems to have been written backwards. That is, explanations are generally attempted only after findings have been surveyed from a vague "let's see ..." approach. This has two negative effects. It makes the reading of summarized empirical findings difficult and tedious, especially in the first two chapters. Of greater concern, this approach seriously weakens the force of the explanations, because they are predominantly *post hoc*. Specific *a priori* predictions are not often set forth for critical test; the arguments lose power since they are not clearly falsifiable. This problem is likely related to the criticism offered next.

It is disturbing that many of the book's ideas are presented with little theoretical and historical background, as though sui generis, when in fact preexisting literature has often addressed a similar or identical view. For example, the discussions about parallels between child and adult phonological properties are quite compatible with Stampe's model of natural phonology, which that author acknowledges in turn as a resurrection of late-19th century phonological theory (e.g., Donegan & Stampe, 1979; Stampe, 1969, 1979). Indeed, Stampe presents an integrated set of specific testable predictions about the phonological properties of child and adult speech, as well as of historical language changes, that could have guided several of the literature searches in Locke's book. Yet Stampe receives only passing mention; likewise, his identified predecessors Sweet, Baudouin, Jespersen, Passy, Hockett, Sievers and others receive scant or no reference. Discussions about the naturalness of phonological properties proceed without clear attribution, and the term natural phonology is even printed in scare quotes, as though newly-coined (p. 141). Similarly, many studies presented as if merely descriptive were actually theoretically motivated, and in directions not altogether dissimilar from that of the book. For example, the treatment of phonological tendencies in speech that has undergone various forms of dissolution (inebriation, dysarthria, aphasia) failed to recognize earlier well-known proponents, notably Ribot (1883), Freud (1953), and Jakobson (1968). A number of other relevant references are also oddly lacking, e.g., Chomsky and Halle (1968), Lieberman (1980); Lieberman et al. (1972), Stark (1980). One would like more evidence of theoretical and historical scholarship, which could have greatly strengthened the thesis of the book by providing a rich source of testable a priori predictions.

There are a number of other, more specific criticisms; I will summarize only a few of the more serious ones here. Discussions about physiological, or neurological, mechanisms that may contribute to the infant's phonetic tendencies are at times confused with anatomical or mechanical factors, and in general are not wholly satisfying. In addition, the sketch in Chapter 2 of a three-stage model for phonological development is interesting but incomplete (age ranges and behavioral markers are unclearly specified); moreover, the description of the first stage is neither phonological nor phonetic. Furthermore, the author notes the striking dissimilarity in the high incidence of /r/ within mature languages vs. its low incidence in infancy and early childhood (during which it is commonly mispronounced when uttered). This fact is a nontrivial challenge to his perspective, yet no serious explanation of the discrepancy was even attempted (there are other such challenges, also under-explained).

Certain peculiarities of style and format need mention. Between-table comparisons of data were made quite difficult, since the format differed widely between tables that were purportedly illustrating the same phonological principles. In at least one case a single table contained some data in percentages, alongside other data presented in raw frequencies (p. 160). The existence of the table formatting discrepancies is perplexing, given the amount of effort that the author obviously spent on interpreting and comparing the data himself! Although the inclusion of a language index is a nice touch, it is frustrating that the book lacks an author index, if one wishes to locate discussion of particular papers. In fact, the quality of the subject index itself is weak, and contains a number of idiosyncratic entries (e.g., Visual pattern imitation in infants, p. 263). Finally, certain stylistic characteristics were distracting, such as idiosyncratic terminology (e.g.,

repertoire vs. nonrepertoire refers to infant babbling sounds that have a high vs. a lower frequency of occurrence, respectively), and liberal and idiosyncratic italicization of quoted passages.

Recommendation

Lest the criticisms appear to overshadow the accomplishments of the book, I must emphasize the service it has provided in ferreting out parallels in phonological and phonetic patterns across a wide array of findings, and in drawing out one view of their implications. The book should serve as an important reference source for specialists in many fields: psycholinguistics, phonology, phonetics, child language, speech science, speech-language pathology, developmental psychology, neuropsychology, even those applying speech science to computer information systems and machine recognition of speech. I concur with the author that it would be additionally useful as a supplement to a main text in courses on language acquisition or phonology, although it is not suitable as a central text itself.

References

Chomsky, N., & Halle, M. (1968). Sound pattern of English. New York: Harper and Row.

Donegan, P. J., & Stampe, D. (1979). The study of natural phonology. In D. A. Dinnsen (Ed.), Current approaches to phonological theory (pp. 126-173). Bloomington, IN: Indiana University Press.

Freud, S. (1953). On aphasia (E. Stengel, trans.). New York: International Universities Press.

Jakobson, R. (1968) Child language, aphasia, and phonological universals (A. R. Keiler, trans.). The Hague: Mouton.

Lieberman, P. (1980). On the development of vowel production in young children. In G. Yeni-Komshian, J. F. Kavanaugh, & C. A. Ferguson (Eds.), Child phonology Vol. 1: Production (pp. 113-142). New York: Academic Press.

Lieberman, P., Harris, K. S., Wolff, P., & Russell, L. H. (1972). Newborn infant cry and nonhuman primate vocalization. Journal of Speech and Hearing Research, 14, 718-727.

Ribot, T. A. (1883). Les maladies de la memoire. Paris: Librarie Germain Bailliere.

Stampe, D. (1969). The acquisition of phonetic representation. Paper presented at the Regional Meeting of the Chicago Linguistic Society, Chicago, IL.

Stampe, D. (1979). A dissertation on natural phonology. New York: Garland Publishing.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. Yeni-Komshian, J. Kavanaugh, & C. A. Ferguson (Eds.), Child phonology Vol. 1: Production (pp. 73-92). New York: Academic Press.

PHONOLOGY AND THE PROBLEMS OF LEARNING TO READ AND WRITE*

Isabelle Y. Liberman+ and Donald Shankweiler+

Abstract. Learning to read and write depends on abilities that are language-related but that go beyond the ordinary abilities required for speaking and listening. Research has shown that the success of learners, whether they are children or adults, is related to the degree to which they are aware of the underlying phonological structure of words. Poor readers are often unable to segment words into their phonological constituents and may have other phonological deficiencies as well. Their difficulties in naming objects and in comprehending sentences, for example, may also stem from a basic problem in the phonological domain.

At the start of formal instruction in reading, the child or adult can speak and understand many words and uncountably many more sentences. Experience tells us, however, that while such command of the language may be necessary for reading, it is not sufficient. But why not? Surely, we must answer that question if we are to understand, and take appropriate action about, the difficulties that so often attend the development of literacy.

Broadly speaking, there are two sets of hypotheses about where the difficulties might lie. One set may be categorized generally as non-language related. Many hypotheses of that kind have been advanced, but perhaps the most widely held (by many clinicians and the lay public, at least) proposes that children who fail have visual perceptual derangements in which they see letters or words wholly or partially backwards. Since the printed word is conveyed to the reader visually, the possibility of some visual defect in the handicapped individual must, of course, be considered. However, we know from the extensive research efforts of many investigators over the years (see Stanovich, 1982, and Vellutino, 1979, for reviews of the evidence) that difficulties in reading are not commonly attributable to perceptual derangements.

---

*To appear in H. Lee Swanson (Special Issue Editor), Memory and learning disabilities: Advances in learning and behavioral disabilities. Adapted from Phonology and the problems of learning to read and write. Topical Issue: Remedial and Special Education, 1985, Vol. 6/6, I. Y. Liberman, Issue Editor.
+Also University of Connecticut.
Acknowledgment. This research was supported in part by grant HD-01994 to Haskins Laboratories from the National Institute of Child Health and Human Development. We are grateful to Alvin M. Liberman for his criticisms of earlier drafts of this paper. We also thank Stephen Crain for his comments on the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]

Our own research and that of others in the field have persuaded us that learning to read and write depends in large part on special language-related skills that go beyond the primary abilities required in producing and understanding speech. But where in language do those skills lie? Early in our research we guessed that many, perhaps most, are in the phonological domain (Liberman, 1971, 1973), and so we put our attention there. For several reasons, that seemed a plausible guess and, therefore, the right place to start: first, because an alphabetic orthography--the kind we must, as a practical matter, be concerned with--represents the phonology, however approximately; second, because the smooth running of the "higher" processes of syntax and semantics presumably depends, at the very least, on the existence of a proper representation in the "lower" domain of phonology (see Liberman, 1983, and Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977, for a discussion of these points). The results of research have, we think, justified our assumptions, providing evidence that characteristics of phonological processing do, indeed, underlie some of the difficulties that poor readers and spellers have. Our aim in this paper is to describe those difficulties and present some of the evidence.

Phonology and Reading the Word

To see what phonology has to do with reading, we must first remind ourselves of what it has to do with language. Perhaps the best way to do that is to imagine what language would be like if there were no phonology. In that case, each word in the language would have to be represented by a signal--for example, a sound--that differed holistically from the signals for all other words. The obvious consequence would be that the number of words could be no larger than the number of holistically different signals a person can efficiently produce and perceive. Of course, we don't know precisely what that number is, but surely it must be small, especially in the case of speech, by comparison with the tens or even hundreds of thousands of words that a language commonly comprises. What a phonology does for us, then, is to provide a basis for constructing a large and expandable set of words--all the words that ever were, are, and will be--out of two or three dozen signal elements. These signal elements, often called phonemes, are themselves represented--though only after complex transformations--by the sounds of speech.

All this is to say that phonology is real--it was not invented by linguists--and, more important, that, whatever else they may be, words are always phonological structures. No matter that the meaning of a word, or its grammatical status, is ambiguous, unknown, or subject to dispute; it is always a string of abstract phonological elements, and, within quite narrow limits, all speakers of the language are in close, if only tacit, agreement about the form of that string. It follows, then, that to have perceived or produced a word, however that may be done, is to have engaged a phonological structure. To misperceive or misproduce a word is to have engaged the wrong phonological structure. We take all of that as given by the very nature of language, as distinguished from other forms of communication such as, for example, pictures.

But why, then, should reading words be difficult in an alphabetic orthography, given that such a transcription represents, if only approximately, the phonological structure that the reader must grasp; and what, as a practical matter, can the teacher do about it? We and our colleagues have offered details in earlier papers (Liberman, 1971, 1973, 1983; Liberman, Liberman, Mat-

tingly, & Shankweiler, 1980; Liberman, Shankweiler, Camp, Blachman, & Werfelman, 1980). Here, it is only appropriate to summarize the argument.

To understand the problem one faces when required to read a word, we must first consider, if only briefly, how the word is perceived when spoken. As we said, the word is formed by a phonological structure, so when the word is perceived, it is this structure that is accessed. But the speaker of the word did not produce the phonological units one at a time, each in its turn--that is to say, he or she did not spell the word out aloud. Rather, the speaker "coarticulated" the phonological units--that is, assigned the consonant we know as 'b,' for example, to the lips, and the vowel we know as 'a,' for example, to a shaping of the tongue, and then produced the two at pretty much the same time. The advantageous result of such coarticulation is that speech proceeds at a satisfactory pace (have you ever tried to understand speech when it was spelled to you, letter by painful letter?), but a further result, and a less advantageous one for the would-be reader, is that there is now, inevitably, no direct correspondence in segmentation between the underlying phonological structure and the sound. Thus, though the word "drag" has four phonological units and, correspondingly, four letters, it has only one pulse of sound, the four elements of the underlying phonological structure having been thoroughly overlapped and merged. How, then, do listeners recover the discrete units of the phonological structure from the seamless sound, thereby making contact with the word as it must be stored in their lexicons?

The long and comprehensive answer has been provided in other papers from our laboratory (see in particular A. M. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; A. M. Liberman & Mattingly, 1985; A. M. Liberman & Studdert-Kennedy, 1978). The short and, for our purposes, sufficient answer is that the phonological segments are recovered from the sound by processes that are deeply built into the aspect of our biology that makes us capable of language. This is to say that in listening to speech, the processes by which we perceive the phonological structure conveyed by speech go on automatically, below the level of conscious awareness. In listening to speech, we are no more consciously aware of the processes by which we arrive at the word than we are consciously aware in vision of the way we use binocular disparity to perceive the relative distance of objects in our field of view.

But reading is different in that it is, in some significant measure, a secondary, less natural, use of language--part discovery, part invention. It follows, then, that even though its processes must at some point make contact with those of the natural and primary system, special skills are required if the proper contact is to be made. We take the point of that contact to be the word, which is, of course, represented in the print by a transcription of the phonological structure. But this transcription will make sense to the child only if he or she understands that it has the same number of units as the word. Only then will the relation between the print and the word be apparent.

Thus, readers can understand, and properly take advantage of the fact, that the printed word <u>drag</u> has four letters, only if they are aware that the spoken word "drag," with which they are presumably already quite familiar, is divisible into four segments. They will probably not know that spontaneously, because, as we have said, the relevant processes of speech perception, which they already command, are automatic and unconscious. And it may be somewhat difficult to teach them what they need to know because, given the overlap of phonological information that characterizes the spoken word, there is no way

to produce the consonant segments in isolation. The teacher can try, of course, to "sound out" the word, but in so doing will necessarily produce a nonsense word comprising four syllables, "duhruhahguh." Such instruction may be better than none at all, but it may not help the child understand why it makes sense to represent the meaningful monosyllable "drag" with four letters. In the next sections, we will offer some of the evidence that shows that novice readers do indeed find it hard to see why, and, further, that their difficulty in this regard is associated with poor reading ability.

Awareness of Basic Phonological Structure

We know that the child's awareness of phonological structure does not happen all at once, but develops gradually over a period of years. Some 12 years ago, we began to examine developmental trends in phonological awareness by testing the ability of young children to segment words into their constituent elements (Liberman, Shankweiler, Fischer, & Carter, 1974). We found that normal preschool children performed rather poorly. We learned, however, as we had suspected, that of the two types of sublexical phonological units, syllables and phonemes, the phonemes presented the greater difficulty. None of the four-year-olds whom we tested could accurately count the number of phonemes in familiar monosyllabic words, though about half managed an accurate count of syllables in multisyllabic words. At the age of five years, a similar pattern emerged: Over half succeeded in the syllable task but less than a fifth could achieve phoneme counting. Only 10% failed the syllable counting task at the end of the first school year, whereas 30% were still failing phoneme counting.

It was clear from these results that awareness of phoneme segments is harder to achieve than awareness of syllable segments, and develops later, if at all. More relevant to our present purposes, it was also apparent that a large number of children may not have attained either level of understanding of linguistic structure, phoneme or syllable, even at the end of a full year in school. We turn now to the evidence that awareness of linguistic structure--an awareness that so many children lack--may be important for the acquisition of reading and spelling.

Awareness of Phonological Structure and Literacy

Much evidence is now available to suggest that awareness of the phonological constituents of words--or as it is sometimes called, metalinguistic awareness--is most germane to the acquisition of literacy. This evidence comes from studies, including some that have been carried out in languages other than English, that have shown that this awareness is predictive of reading success in young children (Alegria, Pignot, & Morais, 1982; Bradley & Bryant, 1983; Liberman, 1973; Lundberg, Olofsson, & Wall, 1980; Mann & Liberman, 1984; deManrique & Gramigna, 1984; Treiman & Baron, 1981). One study, worthy of special mention as one of the most extensive, was carried out in Sweden (Lundberg et al., 1980). Among the many abilities, both related and unrelated to language, considered in that study, the ability to segment words into phonemes was the single most powerful predictor of future reading and spelling skills in a group of children tested at the end of their kindergarten year.

A more modest but similar study from our laboratory (Mann & Liberman, 1984) was a longitudinal comparison of a group of children as kindergarteners and first graders. It had the aim of discovering the best kindergarten pred-

ictors of reading success. The ability to segment words by counting their constituent syllables was selected instead of phoneme counting as the measure of awareness. We knew, given the results of our earlier study, that syllable segmentation ability, unlike phoneme segmentation, was already in place in over half of the children before the first grade; therefore, we considered syllable awareness would be less open to criticism as possibly confounded by reading instruction. Of the 26 children later classified as good readers in the first grade, 85% had "passed" the syllable counting test when they were kindergarteners. In contrast, only 56% of the average readers and 17% of the poor readers had been successful.

In a recent study by our research group (Liberman, Rubin, Duques, & Carlisle, in press), metalinguistic awareness in the phonological domain has also been found to be highly predictive of spelling success. This study, relating the invented spellings (Read, 1971) of kindergarteners to their performance on other language-related tasks, suggests that their proficiency in spelling is more closely tied to phonological awareness than to other aspects of language development. Of the eight language-based tasks administered to this group, three made a difference statistically and accounted for 93% of the variance in invented spelling proficiency. These three unquestionably tapped phonological skills. Listed in descending order of importance, they included a phoneme analysis test patterned after Lundberg et al. (1980); a test of the ability to supply the correct grapheme when phonemes are dictated; and a test of the ability to delete phonemes from spoken words, adapted from the Test of Auditory Analysis Skills (Rosner, 1975). A fourth, a picture naming test, contributed 1% to the variance but did not quite attain significance. It is less obviously phonological in nature, but, as we shall note in a later section, it may be viewed as a subtle indicator of phonological difficulties. The four remaining language-based tasks did not make a difference in the kindergarteners' performance on the invented spelling test. It is notable that although these four tasks all reflect certain aspects of language development, they do not require the degree of awareness of internal phonological word structure that is tapped by the others. Three of these tasks--receptive vocabulary, letter naming/writing, and word repetition--do not include the analytic phonological component at all; the fourth--syllable deletion--taps it at a less abstract level closer to the basic unit of articulation.

These results and the many others that could be cited (Blachman, 1983; Fox & Routh, 1980; Goldstein, 1976; Helfgott, 1976; Zifcak, 1981) certainly suggest that readiness for reading and spelling is related to metalinguistic awareness of the internal structure of words. There is now some evidence that this relationship also implies that phonological awareness may help the child learn to read. This evidence comes from a pair of experiments (Bradley & Bryant, 1983), the first of which looked at the performance of a large number of four- and five-year-olds, none of whom could read, on a metalinguistic task requiring categorization of the "sounds" (phonemic constituents) in words. As in previous studies, high correlations were found between phonological awareness, in this case measured by the sound categorization scores, and the children's reading and spelling scores three years later. The relationship remained strong even when the influence of intellectual level at the time of the initial tests was removed.

However, as the authors themselves correctly point out, simply to show that children's skills in metalinguistic awareness are predictive of their success or failure in reading later on does not by itself prove that the relationship is necessarily a causal one. It is possible, in principle at least, that the measured relationship occurred because both abilities are highly correlated with a third ability and that this unidentified third ability is the controlling factor. In order to get around this problem, the authors carried out a second experiment. This was a training study, using subsamples of the original group, carefully matched for age and IQ, but with initially low scores on phonological judgments. For one subgroup, the training sessions directed the child's attention to shared initial, medial, and final phonemes in consonant-vowel-consonant words. A second group was also taught this information, but in addition was shown how phonemes in the test words could be represented by letters of the alphabet. A third group, a control group, received instruction in semantic classification of the same set of words, but no attention was given to the phonological relationships or the spelling. As an additional control a fourth group received no special training at all. It was found at the end of the project that the children receiving training in phonological categorization were superior to the semantically trained group on standardized tests of reading and spelling, and those trained with alphabetic letters in addition to the phonological training were even more successful (particularly in spelling).

Together, this pair of experiments--combining longitudinal and training procedures--offers the strongest evidence to date of a possible causal link between phonological awareness and reading and writing abilities. At the very least, they support other studies showing that there are methods for training phonological awareness that can be used successfully with young children (Content, Morais, Alegria, & Bertelson, 1982; Olofsson & Lundberg, 1983). Beyond that, they also indicate that this training can have beneficial effects on children's progress in learning to read and spell (see Vellutino, in press, for another phonological training procedure with salutary effects on literacy).

There remains some question, however, concerning the extent to which phonological awareness, which we have seen to be important for reading and spelling success, arises spontaneously, as it were, as part of general cognitive development, or whether, alternatively, it develops only after specific training or as a spinoff effect of reading instruction.

The question as to whether word-related metalinguistic abilities develop spontaneously or must be taught is a crucial one, with obvious implications not only for preschool instruction, but also for the design of literacy teaching programs geared to adults. It was explored in an unusual investigation by a group of Belgian researchers who examined the phonological awareness of illiterate adults in a rural area of Portugal (Morais, Cary, Alegria, & Bertelson, 1979). They found that the illiterate adults could neither delete nor add phonemes at the beginning of nonsense words, whereas others from the same community who had received reading instruction in an adult literacy class succeeded in performing those tasks. The authors concluded that awareness of phoneme segmentation does not develop spontaneously even by adulthood, but arises as a concomitant of reading instruction and experience. A closer look at the results reveals that within the literate group, those who had obtained certificates for passing the course performed significantly better on the measures of phoneme segmentation skill than those who had taken the course but

had not attained the level of proficiency required for a certificate. This kind of variation should not, of course, be ignored. It is entirely plausible that those adults who took the course and did not do well may resemble younger poor readers in other studies: Their failure to develop awareness of phonological structure may have hindered them in learning to read.

Another relevant study is one recently carried out in mainland China with subjects grouped according to whether they had or had not ever been exposed to alphabetic instruction (Read, Ahang, Nie, & Ding, 1984). The results of this study again suggest that reading instruction may be a critical factor in developing phonological awareness. The critical finding is that given a phoneme addition-deletion task (similar to that used with the Portuguese subjects), individuals who at some time in their educational experience had been exposed to pinyin, the official alphabetic spelling system, performed that task very well. In contrast, those whose only literacy training had been in the Chinese logographic characters and who had had no experience with the alphabet did not. Thus, it appears that people who are literate but who have not developed alphabetic literacy may not develop a metalinguistic strategy at the phoneme level.

In view of these findings, we believed that it should prove of value to explore further the cognitive characteristics of adult poor readers. In previous work, we had concentrated on children who were having difficulties learning to read. Now, we proposed to examine the characteristics of adults who, despite years of exposure to alphabetic reading instruction as children, had not achieved full literacy. We were interested in particular to learn whether their performances would be similar to those of younger learners who were having difficulty. We consider a recent study of a community literacy class that was conducted by members of our research group (Liberman, Rubin, Duques, & Carlisle, in press) as only a first step toward that goal, but one that nonetheless provides promising leads.

In a comparison of the reading and spelling of our adult subjects, we found, as would be expected in any comparison of recognition and production measures, that their reading of single real words was better than their spelling of such words. But on nonsense words, for which some explicit reference to the phonological structure is obligatory rather than optional, as it may be in dealing with real words, the advantage of recognition over production was eliminated. The performance of the adults on both reading and spelling of nonsense words was quite poor and virtually identical in quality, bespeaking what seemed to be a serious deficiency in the ability to deal analytically with phonological structure.

The performance of the adult poor readers in another task, one directly measuring language analysis at the phonemic level, lends credence to the hypothesis that they may indeed have such a deficiency. On a very simple phoneme analysis task requiring only that subjects identify the initial, medial, or final sound in words--an exercise commonly encountered in first-grade classrooms, they managed to produce correct responses on only 58% of the items. Moreover, they clearly found the task particularly frustrating and unpleasant. This inability of adults with literacy problems to perform well on tasks requiring explicit understanding of phonological structure has also been found by other investigators (Byrne & Ledez, 1983; Marcel, 1980; Morais et al., 1979; Read & Luyter, 1985).

A recent study of adult prisoners of low literacy (Read & Luyter, 1985) provides strong confirmation of these pilot findings of ours. In their report of this new investigation, the authors note that their subjects remain poor readers despite cognitive maturity, environmental experience with the written language, and adequate general intelligence. The greatest difficulty displayed by these adults is in decoding unfamiliar words and in the segmentation skills that underlie decoding--particularly in tasks that demand awareness of the location of phonemes within a syllable. The subjects are much better at recognizing familiar words and also in tasks that do not require internal phonemic analysis, such as identifying the initial consonant and judging overall similarities in words. The authors remark that whatever the causes of the difficulty--poor educational opportunity and/or motivation--a prominent characteristic now is a disability in decoding new and unfamiliar words and in phonemic segmentation. Moreover, the deficits clearly cannot be attributed to a general maturational lag, for they do not disappear in these adults of adequate intelligence.

Despite much evidence of the kind we have been considering here, there remains a question as to whether the deficiency may not in fact be necessarily phonological, or even linguistic, but rather attributable to a deficiency in general analytic ability (Wolford & Fowler, 1983). This question is addressed directly, and, in our view, very convincingly, in a recent study by the Brussels group of experimenters. They have recently shown (Morais, Cluytens, & Alegria, 1984) that poor readers--in this case, children aged six to nine with severe reading disability--were poorer than normal readers in segmenting words into their constituent parts, but performed as well as normal readers in a similar task that required them to deal not with words but with musical tone sequences. Thus, evidently the deficiency that the poor readers were exhibiting was not due to a general analytic disability, but was rather specifically language-related and, more than that, specifically phonological in nature.

The possible presence in poor readers of a general analytic deficiency rather than a deficiency specifically in the phonological realm was a question also addressed in yet another recent study (Pratt, 1985). There two complementary experiments were carried out--one with good and poor readers in adult education classes and the other with good and poor readers in the third grade. Both reader groups in each case were given linguistic awareness tasks and a nonspeech control task identical in format to one of the linguistic tasks. Significant differences between the good and poor readers at both levels were found on all three linguistic awareness measures but not on the nonspeech control task.

Thus, it appears again that the deficiency the poor readers were exhibiting was not due to some general analytic disability, but was, instead, specifically language-related and, more than that, specifically phonological in nature.

As we have seen, there is now a wealth of evidence pointing to metalinguistic deficiencies in the phonological domain in individuals of various ages, languages, and cultural backgrounds, who have difficulty in attaining literacy. We suggest that perhaps it would be reasonable now to consider seriously the possibility that the deficiency in these individuals who are resistant to ordinary methods of literacy instruction may not be limited to metalinguistic awareness, but may reflect a more general deficiency in the phonological domain. Some of the evidence for this conjecture will be discussed in the next two sections.

Phonology and Naming

We now turn to consider the significance of the well-known fact that children who are poor readers often have some degree of difficulty in producing the names of things. At first blush, this would appear to be a problem completely separate from their difficulties in reading. But, in our view, the failures in calling up the appropriate name of an object and the failures in identifying words in print may both relate in some degree to the poor readers' difficulties with language at the level of the phonology.

Several investigators have found that errors in naming are characteristic of children with reading disability (Denckla & Rudel, 1976; Jansky & de Hirsch, 1973; Katz, in press; Mattis, French, & Rapin, 1975; Wolf, 1981). The existence of a naming problem can be demonstrated by a picture naming test of the sort that is commonly used in testing aphasic patients. The data we will discuss here were obtained using an adaptation of the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1976), in which the subject is presented with pictured objects one at a time and is required to name each item as it appears.

The fact that poor readers tend to misname things could lead one to infer that the problem is semantic. But, as we shall see, this may be a wrong inference. The first step toward a correct analysis of the poor reader's naming difficulties is to recognize that there are several different aspects to the naming task. First, the perceiver has to apprehend the object in perception. The object must be recognized for what it is. Then a search of the internal lexicon must be carried out to find the word that best names the object. Finally, the word must be articulated in overt speech. An error can arise at any stage from perceptual apprehension to phonetic output. Thus, an error in naming does not automatically reveal its source, which can only be discovered by further analysis.

The experiments needed to pinpoint the source of mistakes in naming have rarely been carried out. Katz's (in press) study is noteworthy in this regard. Words selected for the study were pictured items from the Boston Naming Test that were considered appropriate for children aged 8-10. High-frequency and low-frequency words were equally represented in this revised version of the test.

In tabulating the results, Katz noted the relationship between each naming error and the target word (i.e., the word judged to be the best name for the object depicted). He showed that although the poor readers produced more incorrect names than the good readers, their responses were not arbitrary. Indeed, they often resembled closely the phonological structure of the correct word. For example, when the picture presented was of a globe, one child's response was to produce the nonword, <u>gloave</u>, which, though incorrect, is identical to the target word except in the last phonological segment. Such an error is consistent with the hypothesis that the child has identified the object in question, but has difficulty producing the word.

In other cases, the child produced a real word in response to the test picture. Again, the response often bore a close phonological resemblance to the target word phonologically. Thus a frequent response to the picture of a volcano was the word <u>tornado</u>--quite different in meaning but with the same number of syllables, an identical stress pattern, and similar vowel

constituents. Without further tests, however, the interpretation of such a response would be ambiguous. Katz resolved these ambiguities by questioning the child. When, in this instance, the subject was subsequently quizzed about the characteristics of the pictured object, he correctly described a volcano and not a tornado. Thus, it was clear that the child was quite aware of the meaning of the object. Many other cases in which an ambiguous response was produced were resolved similarly: It often turned out that the child's problem had to do not with meaning, but with the phonological structure of the target word. Thus, whether the poor readers' responses were nonwords, as in the first example, or incorrect real words, as in the second example, the source of the error was often phonological.

Further indications that phonology and not semantics may have been at the basis of these poor readers' naming errors are provided by the results of a test of identification of pictured objects in which the previous procedure was reversed. In this reversed procedure, the examiner produced the name and the child had to select the one picture from a set of eight that best depicted the meaning of the word. Each item that had previously been misnamed on the naming test was subsequently tested for recognition in this manner. In most cases, correct retrieval was demonstrated. Thus, it was apparent that the poor readers had acquired internal lexical representations of most of the objects whose names they could not produce accurately. As Katz (in press) points out, distorted production of the word for an item that has been correctly identified could stem either from an incomplete specification of the phonological word in the lexicon, or from deficient retrieval and processing of the stored phonological information. Which of these possibilities is correct is not relevant to the question at issue here. What is relevant is that, in either case, the source of the poor readers' difficulty had to do with the phonologic aspect of words and not with their meanings.

Phonology and Sentence Comprehension

Having seen that deficiencies in the phonological domain may be responsible for difficulties in reading words, and also for some of the well-known problems of naming, we turn to the role of phonological abilities in sentence comprehension. Recent investigations have noted that poor readers frequently have difficulties understanding complex sentences, not only in reading but also in speech (Byrne, 1981; Vogel, 1975). Our principal task in this section is to say why one would suppose that the deficit that underlies poor readers' difficulties in sentence understanding is phonologic, and how we have gone about testing this idea.

We begin by making three points: First, understanding sentences requires short-term memory. Second, short-term memory depends on the ability to exploit phonological structure. Third, young children who are poor readers are known to have special limitations in short-term memory and deficiencies in the use of phonological structure. We will take up each of these points in turn and attempt to show the connections between them. First, we will discuss how short-term memory is relevant for comprehension, then we will suggest how the short-term memory system depends on phonological structures, and finally we will introduce evidence that the comprehension problems of poor readers stem not from lack of syntactic abilities but from weaknesses in the phonologic system.

It has been suggested that short-term storage must play a central role in the operation of the syntactic and semantic processors because ascriptions of syntactic structure and propositional content must be based on briefly holding sequences of words in memory (Liberman, Mattingly, & Turvey, 1972). Thus, verbal short-term memory is needed for processing connected discourse, whether it is apprehended through the medium of the printed page or by speech. Although use of short-term memory is not unique to reading, we will argue that reading may place special demands on this system.

The hypothesis regarding need for short-term memory might seem to be weakened by recent data from several sources indicating that the processes supporting sentence comprehension are to a considerable extent performed "on line" (e.g., Frazier & Fodor, 1978; Frazier & Rayner, 1982). Partly in response to such findings, most recent current conceptions of sentence parsing mechanisms have the parser operating on small chunks of the text (groups of two or three words). In our view, these developments actually strengthen the argument that short-term memory is essential to ongoing language processing. It is precisely because this memory system has such a limited capacity for retention of the verbatim record that fast-acting processing routines must have evolved (Crain & Shankweiler, in press). There is much evidence that the temporary memory system, on which the processing of connected language depends, briefly preserves the phonology and its phonetic derivatives--short-term memory is thus said to depend on an internal phonetic code (Conrad, 1964, 1972; Crowder, 1978).

In relating this information about memory to the performance of beginning readers, it is significant, first, that the memory deficits of young children who are poor readers appear to be limited, by and large, to the linguistic domain. For example, we have found that they have no more difficulty than good readers with memory for faces, nonsense designs, and other stimuli not amenable to verbal labeling (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, 1982). In addition, there is reason to believe that poor young readers are specifically deficient in use of the short-term memory code. Thus, it has been found that poor readers in the early elementary grades, who perform poorly also on tests of immediate recall, do not code the phonetic properties of words as fully as good readers (Brady, Shankweiler, & Mann, 1983; Liberman et al., 1977; Olson, Davidson, Kliegl, & Davies, 1984; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979).

Considerable evidence already exists pointing to a connection between poor readers' difficulties in remembering sequences of spoken words (and other materials that can be coded as words) and their failure to exploit phonological structure as a vehicle for short-term retention (Mann, Liberman, & Shankweiler, 1980). The suggestion has also been made (Byrne, 1981; Mann et al., 1980; Shankweiler et al., 1979; Vellutino, 1979) that short-term memory limitations might account as well for the problems poor readers sometimes display clinically in oral sentence comprehension. This possibility was strengthened by the finding that poor readers are worse than good readers not only in recall of arbitrary strings of words, but also in recall of both meaningful and meaningless (but syntactically accurate) sentences (Mann et al., 1980).

Until a recent study by Mann, Shankweiler, and Smith (1984), however, no experiment had expressly addressed the question of whether the sentence comprehension problems of poor readers might not be to some degree phonologic in nature, rather than syntactic. The test of syntactic competence selected

to make this determination tapped the subject's understanding of relative clauses. The relative clause, which allows the embedding of sentences within one another, was chosen because it is a device of central importance to grammatical function. Syntactically complex, it is apt to be misinterpreted by young children (Tavakolian, 1981) and also by older persons with language disorders (Caramazza & Zurif, 1976).

Good and poor readers in the third grade were tested for comprehension of four different orally presented relative clause structures. In constructing the test sentences, account was taken of the grammatical fact that a relative clause may attach either to a subject noun phrase or to a direct-object noun phrase, and, further, that the relative pronoun that substitutes for the missing noun phrase (in the relative clause) can take either the subject role or the direct-object role.

Comprehension of the tape-recorded sentences was tested by the children's manipulation of toy animals. Rote recall for the sentences was also tested, but on a later day; the children listened to the recordings again and were asked to repeat each sentence as accurately as possible. The pattern of errors for good and poor readers in comprehension and recall for each type of relative-clause sentence was then examined. One way an error of sentence interpretation can arise is from simplification of the structure of a sentence containing a relative clause. For example, the sentence might be interpreted as having two main clauses joined by and rather than having a relative clause modifying a noun phrase. Such an erroneous parsing of a sentence containing an object-relative clause, as in the example, "The dog stood on the turtle that chased the sheep," would result in a response by the child in which the dog stands on the turtle and chases the sheep. If it were found that poor readers made chiefly this kind of error, it could be taken to imply that their grammar is less differentiated than that of normal adults and more mature children of their own age. Such a finding would constitute evidence of a primary deficiency in syntactic competence. But, in the event, that is not what happened.

Turning to the results of the test of comprehension, we consider first the errors for each of the four sentence types, separately for good and poor readers. It was found that the poor readers made consistently more errors than the good readers. It was expected, on the basis of past research on language acquisition (Tavakolian, 1981), that there would also be differences in difficulty among the sentence types, and, in fact, such differences were found even in children as old as these (8-10 years). But when the four sentence types were ranked in order of difficulty for good and poor readers separately, the ordering was found to be the same for both groups. The poor readers were generally worse than the good readers in comprehension of relative clause sentences, but within this broad class, they were affected by syntactic variations in the same way as the good readers. The results give no evidence, then, that the poor readers were deficient on any facet of the grammar pertaining to the interpretation of these relative clause sentences. The competence they displayed in this regard was essentially like that of the good readers. A similar result was obtained in a second experiment on interpretation of reflexive pronouns that employed the same subjects (Shankweiler, Smith, & Mann, 1984).

We must account, however, for the other major finding of the study: The poor readers' performance, though similar in pattern, was not equivalent in proficiency to that of good readers in comprehension of any of the four relative clause structures. The best clue we have as to why the poor readers were less accurate is given by comparing their performance on the test of rote recall, where it was found that the poor readers also made significantly more errors. Again, the differences between the groups did not favor one type of sentence more than another. When the recall scores and the comprehension scores on individual subjects are compared statistically, a significant degree of correlation is found. These results are also in complete agreement with recall findings obtained earlier (Mann et al., 1980) with comparable groups of good and poor readers. They fit well with much earlier work that indicates, as we have seen, that poor readers perform consistently more poorly than good readers on a variety of tests of verbal short-term memory. Thus the failure of the poor readers to do as well as the good readers on the test of sentence comprehension is probably a reflection, at least in part, of verbal short-term memory deficiencies in the poor reader group.

Although these studies do not totally resolve the question of whether the poor readers have a deficit in syntactic competence as such, there is nothing in the findings that would specifically indicate such a deficit. Instead, the findings suggest that our disabled readers have acquired the grammar they need for understanding these complex sentences, though they do not always interpret them correctly. When they deviate from good readers, it would appear to be because they cannot remember the words and their order of occurrence as well. Thus the findings we have to date support the claim that the poor readers' difficulties in comprehension may ultimately stem from failure to exploit the phonological structure in short-term memory. Therefore, we would suppose that the difficulties in understanding sentences, like the difficulties in reading words and naming objects, are at root phonological.

The phonological deficiencies we have uncovered in poor readers' performance on tasks involving spoken language have definite consequences for reading and it is to reading comprehension itself that we now turn. It is important to appreciate that the problems that poor readers characteristically have in comprehension of text stem in large part from their slow and inaccurate word decoding skills. Because short-term memory is, for everyone, both fleeting and limited in capacity, the rate at which material is read into short-term memory is critical. Perfetti and his colleagues (Perfetti & Hogaboam, 1975) have suggested that poor readers cannot use their short-term memory efficiently because of the "bottleneck" created by slow word recognition. Thus reading sentences with comprehension would be hampered, even if all the component words were identified correctly, but too slowly to be processed efficiently. The problem is even more serious, however, than we have indicated so far. Poor readers, as we have seen, have not just the normal limitations of short-term memory; their short-term memory spans are abnormally curtailed. Therefore, poor readers' problems in reading complex sentences may be especially acute.

The point that we would add to this account of the bottleneck hypothesis is that, in view of the findings of Mann et al. (1984), we do not have to invoke a syntactic deficit in order to account for problems in reading sentences. We see that a low-level deficit in use of the orthography to gain access to word representations may have major repercussions on the higher-level syntactic and semantic processes required for text comprehension, especially

when compounded by a short-term memory problem. Our research leads us to believe that reading comprehension difficulties may reflect processing limitations originating in the phonology, and not necessarily absence or malformation of the higher level structures of the sentence grammar.

Summary and Conclusions

In our research we have sought to identify the language-related sources of difficulty in learning to read and write. To this end, we have explored the difficulties of poor readers in reading words, in naming, and in sentence comprehension. First, we discussed evidence suggesting that it is difficult for the beginning reader to grasp that words have parts: phonemes, syllables, morphemes. A language user does not need to be aware of what the parts are in order to speak and understand speech because the built-in speech apparatus processes them automatically. But to learn to use an alphabet, to read and to spell, the learner needs to become aware of the parts to make the connection between speech and writing. Awareness of sublexical structure draws upon a set of phonological (or, more accurately, morphophonological abilities [Liberman, Liberman, Mattingly, & Shankweiler, 1980]). Possession of these abilities distinguishes people who are good readers and spellers from those who are less skilled. Though native abilities may account to a considerable degree for the differences, experience in reading and writing also plays a significant role.

Poor readers not only have problems in identifying printed words, they also frequently have problems finding the most appropriate words for things in speaking. By quizzing poor readers about the objects they misname, it has been learned that the source of the naming error is not always a semantic confusion. Frequently, the source of the problem is not having ready access to the mental structures that store information about the phonological properties of particular words in the vocabulary (Katz, in press).

In the last section of the paper we showed that difficulties in the phonologic domain are sufficient to cause problems in sentence understanding. In order to process complex sentences accurately, one needs to have the ability to retain the words of the sentence and their order, briefly, while the information is processed through the several levels from sound to meaning. Poor readers do not remember ordered series of linguistic items (words and objects that can readily be coded as words) as well as good readers. Their special-purpose phonetic working-memory system is deficient. This is probably not a general cognitive deficit, since nonlinguistic memory tests do not distinguish poor readers from good readers. The processing limitation, which is apparently specific to systems that support language use, can affect comprehension when the sentence structure is complex even though the basic grammar is, to the best of our knowledge, intact. It can also lead to severe difficulties in the comprehension of printed text because short-term memory function is hobbled by slow and inaccurate word recognition.

We have identified three problems of the poor reader--difficulty in becoming aware of sublexical structure for the purpose of developing word-recognition strategies, unreliable access to the phonological representations in the internal lexicon for naming objects and for performing metalinguistic tasks involving phonological properties of words, and finally, the deficient use of phonetic properties as a basis for the short-term working memory operations that underlie the processing of connected language in any

form. We cannot fail to notice that all of these are deficits in "lower level" abilities. It is an important task for future research to determine how these abilities, each of which involves the phonological component of the language apparatus, are related in development and pathology.

There is now much evidence that metalinguistic abilities in the phonological domain can be taught at all ages with significant success. Moreover, there is increasing evidence that such phonological instruction has beneficial effects on proficiency in reading words. We know relatively little about the role of instruction in developing and maintaining or expanding the phonetic short-term memory system required for sentence comprehension. But whether or not phonetic memory function can be improved by instruction, we know that pressure on short-term memory is reduced as reading strategies become more efficient. Thus, fostering phonological development in the beginning reader may serve to improve not only the reading of words, but also the comprehension of sentences. Various ways to promote phonological development have been outlined elsewhere (Bradley & Bryant, 1983; Liberman, Shankweiler, Camp, Blachman, & Werfelman, 1980; Olofsson & Lundberg, 1983). However, the creative teacher who understands the basic problems the child faces in learning to read and write will have no trouble devising other, equally appropriate, techniques.

References

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. Memory & Cognition, 10, 451-456.

Blachman, B. (1983). Are we assessing the linguistic factors critical in early reading? Annals of Dyslexia, 33, 91-109.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read--a causal connection. Nature, 301, 419-421.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, 35, 345-367.

Byrne, B. (1981). Deficient syntactic control in poor readers: Is a weak phonetic memory code responsible? Applied Psycholinguistics, 2, 201-212.

Byrne, B., & Ledez, J. (1983). Phonological awareness in reading-disabled adults. Australian Journal of Psychology, 35, 185-197.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. Brain and Language, 3, 572-582.

Conrad, R. (1964). Acoustic confusions in immediate memory. British Journal of Psychology, 55, 75-84.

Conrad, R. (1972). Speech and reading. In J. Kavanagh & I. Mattingly (Eds.), Language by ear and by eye: The relationships between speech and reading (pp. 205-240). Cambridge, MA: MIT Press.

Content, A., Morais, J., Alegria, J., & Bertelson, P. (1982). Accelerating the development of phonetic segmentation skills in kindergarteners. Cahiers de Psychologie Cognitive, 2, 259-269.

Crain, S., & Shankweiler, D. (in press). Syntactic complexity and reading acquisition. In A. Davison, G. Green, & G. Herman (Eds.), Critical approaches to readability: Theoretical bases of linguistic complexity. Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1978). Language and memory. In J. F. Kavanagh & W. Strange (Eds.), Speech in the laboratory, school, and clinic (pp. 331-375). Cambridge, MA: MIT Press.

Denckla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. Brain and Language, 3, 1-15.

Fox, B., & Routh, D. K. (1980). Phonetic analysis and severe reading disability in children. Journal of Psycholinguistic Research, 9, 115-119.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. Cognition, 6, 291-325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14, 178-210.

Goldstein, D. M. (1976). Cognitive-linguistic functioning and learning to read in preschoolers. Journal of Educational Psychology, 68, 680-688.

Helfgott, J. (1976). Phoneme segmentation and blending skills of kindergarten children: Implications for beginning reading acquisition. Contemporary Educational Psychology, 1, 157-169.

Jansky, J, & de Hirsch, K. (1973). Preventing reading failure. New York: Harper and Row.

Kaplan, E., Goodglass, H., & Weintraub, S. (1976). Boston naming test. Boston: Boston University School of Medicine.

Katz, R. B. (in press). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. Cognition.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. Journal of Experimental Child Psychology, 32, 474-484.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

Liberman, A. M., Mattingly, I. G., & Turvey, M. (1972). Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes and human memory (pp. 307-334). Washington, DC: Winston and Sons.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII: Perception (pp. 143-178). New York: Springer-Verlag.

Liberman, I. Y. (1971). Basic research in speech and lateralization of language: Some implications for reading disability. Bulletin of the Orton Society, 21, 71-87.

Liberman, I. Y. (1973). Segmentation of the spoken word and reading acquisition. Bulletin of the Orton Society, 23, 65-77.

Liberman, I. Y. (1983). A language-oriented view of reading and its disabilities. In H. Myklebust (Ed.), Progress in learning disabilities (Vol. 5, pp. 81-101). New York: Grune & Stratton.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading, and dyslexia (pp. 137-153). Austin, TX: Pro-Ed.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. (1982). Children's memory for recurring linguistic and non-linguistic material in relation to reading ability. Cortex, 18, 367-375.

Liberman, I. Y., Rubin, H., Duques, S. L., & Carlisle, J. (in press). Linguistic skills and spelling proficiency in kindergarteners and adult poor spellers. In J. Kavanagh & D. Gray (Eds.), Dyslexia: Biology and behavior. Parkton, MD: York Press, Inc.

Liberman, I. Y., Shankweiler, D., Camp, L., Blachman, B., & Werfelman, M.

(1980). Steps toward literacy. In P. Levinson & C. Sloan (Eds.), Auditory processing and language: Clinical and research perspectives (pp. 189-215). New York: Grune & Stratton.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. Journal of Experimental Child Psychology, 18, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conferences (pp. 207-225). Hillsdale, NJ: Erlbaum.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. Scandinavian Journal of Psychology, 21, 159-173.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory: Can they presage early reading problems? Journal of Learning Disabilities, 17, 592-599.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. Journal of Child Language, 11, 627-643.

deManrique, A. M. B., & Gramigna, S. (1984). La segmentacion fonologica y silabica en ninos de preescolar y primer grado. Lectura y Vida, 5, 4-13.

Marcel, A. (1980). Phonological awareness and phonological representation: Investigation of a specific spelling problem. In U. Frith (Ed.), Cognitive processes in spelling (pp. 373-403). London: Academic Press.

Mattis, S., French, J. H., & Rapin, I. (1975). Dyslexia in children and young adults: Three independent neuropsychological syndromes. Developmental Medicine and Child Neurology, 17, 150-163.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? Cognition, 7, 323-331.

Morais, J., Cluytens, M., & Alegria, J. (1984). Segmentation abilities of dyslexics and normal readers. Perceptual and Motor Skills, 58, 221-222.

Olofsson, A., & Lundberg, I. (1983). Can phonemic awareness be trained in kindergarten? Scandinavian Journal of Psychology, 24, 35-44.

Olson, R. K., Davidson, B. J., Kliegl, R., & Davies, S. E. (1984). Development of phonetic memory in disabled and normal readers. Journal of Experimental Psychology, 37, 187-206.

Perfetti, C. A., & Hogaboam, T. (1975). The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 67, 461-469.

Pratt, A. (1985). The relationship of linguistic awareness to reading skill in children and adults. Unpublished doctoral disseration, University of Rhode Island.

Read, C. (1971). Pre-school children's knowledge of English phonology. Harvard Educational Review, 41, 1-34.

Read, C., Ahang, Y., Nie, H., & Ding, B. (1984). The ability to manipulate speech sounds depends on knowing alphabetic spelling. Paper presented at the International Congress of Psychology, Acapulco, Mexico.

Read, C., & Luyter, L. (1985). Reading and spelling skills in adults of low literacy. In I. Y. Liberman (Ed.), Topical issue: Phonology and the problems of learning to read and write. Remedial and Special Education, 6/6.

Rosner, J. (1975). Helping children overcome learning difficulties. New York: Walker.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 5, 531-545.

Shankweiler, D., Smith, S. T., & Mann, V. A. (1984). Repetition and comprehension of spoken sentences by reading-disabled children. Brain and Language, 23, 241-257.

Stanovich, K. E. (1982). Individual differences in the cognitive processes of reading: 1. Word decoding. Journal of Learning Disabilities, 15, 449-512.

Tavakolian, S. L. (1981). The conjoined-clause analysis of relative clauses. In S. Tavakolian (Ed.), Language acquisition and linguistic theory (pp. 167-187). Cambridge, MA: MIT Press.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (Vol. 3). New York: Academic Press.

Vellutino, F. R. (1979). Dyslexia: Theory and research. Cambridge, MA: MIT Press.

Vellutino, F. R. (in press). Phonological coding: Phoneme segmentation and code acquisition in poor and normal readers. In J. Kavanagh, D. Gray, & D. Pearl (Eds.), Dyslexia: Biology and behavior. Parkton, MD: York Press, Inc.

Vogel, S. A. (1975). Syntactic abilities in normal and dyslexic children. Baltimore, MD: University Park Press.

Wolf, M. (1981). The word-retrieval process and reading in children and aphasics. In K. Nelson (Ed.), Children's language (Vol. 3). New York: Gardner.

Wolford, G., & Fowler, C. A. (1983). Perception and use of information by good and poor readers. In T. Tighe & B. Shepp (Eds.), Perception, cognition, and development: Interactional analyses (pp. 267-292). Hillsdale, NJ: Erlbaum.

Zifcak, M. (1981). Phonological awareness and reading acquisition. Contemporary Educational Psychology, 6, 117-126.

PHONOLOGICAL DEFICIENCIES IN CHILDREN WITH READING DISABILITY: EVIDENCE FROM AN OBJECT-NAMING TASK*

Robert B. Katz†

Abstract. Research indicates that children with reading disability have problems both in naming objects and in performing certain tasks that require phonological processing or phonological awareness. The present study explored the possibility that these problems are related: Poor readers may have object-naming deficits as a consequence of phonological deficiencies in establishing complete representations in long-term memory and in processing these representations. This hypothesis was supported in an initial experiment that required children to name pictured objects. The poor readers were less accurate than the good readers in labeling the objects. Their difficulty was particularly marked on objects with low frequency names and those with polysyllabic names, these being, presumably, more difficult to represent and to process accurately than frequent and short names. Moreover, the incorrect responses bore a phonetic resemblance to the correct object names. In a second experiment, the poor readers had difficulty making decisions based on the length of object names, even when it could be established that they knew the names. This suggests that they lack explicit awareness of the correspondence between the units of phonological representations and the units of speech. Since there is evidence that this awareness is important for learning to read well, the findings of this experiment and the first experiment support the hypothesis that the difficulties of poor readers reflect common stages in the processes that underlie reading and naming.

Errors in naming objects are characteristic of children with reading disability (Denckla & Rudel, 1976; Jansky & deHirsch, 1973; Mattis, French, & Rapin, 1975; Wolf, 1981). On tests of naming, it is usual for such children

---

*Cognition, in press.
†Now with the University of Miami and the Veterans Administration Medical Center, Miami.
Acknowledgment. This article is based on a doctoral dissertation submitted to the University of Connecticut. I would like to thank Donald Shankweiler for valuable advice during the course of this project. Also, thanks go to Carol Fowler, Leonard Katz, Isabelle Liberman, Ignatius Mattingly, and Michael Studdert-Kennedy. Finally, I would like to thank Frances Klein, supervisor of reading for the East Hartford, Connecticut, public school system, and the children and teachers of Hockanum School for their cooperation in this study. The research reported here was supported by NICHD grant HD-01994 to Haskins Laboratories and by a dissertation fellowship from the University of Connecticut Research Foundation.

to name fewer of a set of pictured objects correctly than normal readers of the same age. In fact, the co-occurrence of naming and reading problems is found even among poor readers who score normally on intelligence tests and who have no obvious difficulties with spoken language. Although the occurrence of naming deficits in poor readers has been recognized for some time, the reasons they occur, and the relations they may have to reading problems, are matters that research has scarcely addressed. The present study provides new data that address these questions. The naming performance of reading-disabled children was investigated in the context of the children's other language-related problems on the expectation that an interpretable pattern of deficits could be elicited. The findings lead to a consideration of the possibility that phonological deficiencies might underlie both the children's naming deficits and their reading difficulties.

Some preliminary remarks on the naming act will indicate the rationale for the method of the present study. The starting point for naming is an object in the world and the endpoint is the production of a word that is the best label for a given object. A number of mental processes intervene. The first requirement is registration of the object in perception. Since the name of an object is not inherent in the object itself, a phonological representation of the name must then be located by a search of long-term memory. There is reason to believe (Labov, 1973; Miller, 1978) that the search may be influenced by stored semantic information, such as knowledge of the use for which the object is employed. Further, once the representation is located, it must be processed (i.e., given a phonetic interpretation) in order to articulate the object's name.

Thus, three broad classes of processes have been acknowledged in models of naming (Caramazza & Berndt, 1978; Goodglass, 1980; Wolf, 1981): perceptual, semantic, and phonological. A deficiency in any one of these could lead to failure in naming. A perceptual or a semantic deficiency could prevent an object from being recognized and identified. In contrast, deficiency in processing a phonological representation could prevent the individual from generating the accepted name even though the appropriate phonological representation had been located. Thus, naming deficits can occur in a number of ways. The occurrence of a naming error does not reveal its source without further analysis.

The aim of the present study was to confirm the existence of naming deficits in poor readers and to probe specifically for the phonologically-related deficiencies that may underlie them. This approach was adopted because a variety of evidence indicates that poor readers have weaknesses in the phonological domain. Their problems are evident in several laboratory tasks. Poor readers are less aware than good readers of the phonetic segments of spoken language (Liberman, Shankweiler, Fischer, & Carter, 1974) and less able to extract the phonetic information from speech stimuli degraded by noise (Brady, Shankweiler, & Mann, 1983). On short-term memory tasks, poor readers are less able than good readers to exploit phonetic properties in retention of the items and their serial order (Katz, Shankweiler, & Liberman, 1981; Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Mann, Liberman, & Shankweiler, 1980; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979). On long-term memory tasks, problems have also been found in poor readers' ability to learn new words (Nelson & Warrington, 1980). There are reasons, then, for suspecting that a deficiency in the phonological aspects of object naming could underlie the deficits of poor

readers on naming tasks. Although this possibility has been raised in earlier discussions of reading disability (Denkla & Rudel, 1976; Wolf, 1979, 1981), it has never been investigated systematically.

In earlier research (Wolf, 1979), semantic similarities between errors in naming and the target items have often been noted (e.g., "hose" for "nozzle," or "Eskimo house" for "igloo"). Such so-called "semantic" errors can, of course, result from a misidentification of the object.[1] But, alternatively, semantic errors may be a consequence of the putative phonological deficiencies. These deficiencies could make it impossible for the child to use an existing phonological representation as the basis for correctly articulating the object name. In such cases, children may be compelled to substitute one or more words that are better represented or that can be more easily processed. It may sometimes happen that when a semantically-related word is substituted, it will also be related phonetically to the correct response (e.g., "seashell" for "seahorse"). This is found to be true of semantic errors that occasionally occur in normal spontaneous utterances (Fay & Cutler, 1977). It is easy to imagine a parallel in mistakes of naming. The influence of the "correct" phonological representation of the object name on the error may be revealed whenever a phonetic resemblance is present. Following this line of reasoning, the effect of phonological deficiencies can be assessed, at least in part, by comparing the phonetic similarity of the erroneous response to the target item. In contrast, attempting to classify the errors into categories, such as "phonetic" versus "semantic," would not be appropriate, since phonological deficiencies could conceivably result in errors of both types.

The hypothesis that the naming deficits of children who are poor readers are often due to phonological deficiencies can thus provide a principled account of naming errors. Moreover, this proposal has a major advantage over alternative accounts: it can rationalize the occurrence of naming deficits in conjunction with reading problems.[2] The same phonological deficiencies could lead to problems in both naming and reading, because each function depends critically on the efficient operation of certain phonological abilities. In reading, one can argue that the representations of words are accessed via the phonology that is reflected in the orthography of printed words (Liberman, Liberman, Mattingly, & Shankweiler, 1980). Once a phonological representation is accessed, the phonetic form of the word can then be derived. In naming an object, the way in which a phonological representation is accessed must be entirely different. Since the object itself does not inherently represent the phonology of the language, the representation is accessed by using perceptual and semantic information. But after accessing the representation, the child must use it as the basis for generating the phonetic code to be articulated, just as would be the case in reading. If the child's phonological representations are incomplete, or if his/her processing of the representations is inefficient, then deficits in both reading and naming would be expected to occur as a consequence. Thus, the co-occurrence of reading and naming disorders can be rationalized by proposing that both are based on the same phonological deficiencies.

Two experiments were conducted to examine the hypothesis that phonological deficiencies contribute to the object-naming deficits of poor readers. In the first experiment, children who varied in reading ability were required to name pictured objects in order to confirm the existence of naming deficits in the poor readers. Evidence that the failure to name objects

correctly was due to phonological deficiencies was sought by analyzing the erroneous responses and by analyzing the characteristics of the object names that were produced incorrectly. In a second experiment, the same children were compared on their ability to make metalinguistic decisions based on the names of pictured objects. The children were tested on two metalinguistic tasks that differed in the kinds of phonological attributes that were relevant to successful execution. Each task required that the necessary phonological attributes be adequately represented and that the subject have conscious access to these attributes.

Experiment 1

The purpose of the first experiment was to confirm the existence of naming deficits in poor readers and to determine the basis of any deficits that might be found. Accordingly, children who differed in reading ability were asked to name line drawings of objects as quickly as possible. By stressing speed of response, it was expected that the children's naming ability would be taxed, thus eliciting errors. On those trials in which the correct name was not produced, further testing was done with the aim of assessing possible tacit knowledge of the name and of assessing familiarity with the pictured object. Then, a phonetic prompt to the correct response was provided, consisting of the initial consonant(s) and vowel of the target word. A post-test was conducted to determine whether the names of the objects were actually represented in the children's lexicons. On this test, the children were presented with sets of pictured objects, most of which had been presented earlier on the naming test. The task was to point to the objects as they were named by the experimenter. The recognition post-test was necessary in order to exclude the possibility that the poor readers could name fewer objects merely because they have smaller vocabularies than the better readers.

Evidence that the failure to name objects correctly can be attributed to phonological deficiencies was obtained in three ways. First, the degree of phonetic relationship between the erroneous response and the correct object name was analyzed. It was expected that phonological deficiencies would lead to errors that phonetically resemble the target names. This would be true of both the good and the poor readers, but, whereas the poor readers were expected to make many errors of this kind, the good readers were expected to err on the few object names that either are not fully represented or are not processed effectively. Second, the children were tested on their awareness of the length of the names of objects that were labeled incorrectly. It was expected that on this metalinguistic test all the children could provide evidence that certain gross phonological characteristics of most of the words, such as their length, were represented even though processing deficiencies may have prevented the production of the words. Third, the effect of word frequency and word length on object naming was examined. It was expected that objects with names that are low frequency words would tend to be labeled incorrectly since the names, having been encountered infrequently, would be incompletely represented. Objects with long names may also be difficult to label, since longer words require that more phonological information be represented and processed. Due to their general phonological deficiencies, it was expected that the poor readers would make disproportionately more errors than the good readers both on low frequency words and on long words.

Method

Subjects

The subjects were children selected from three third-grade classes in a suburban Connecticut public school. All those for whom parental permission was obtained were eligible for testing. Of the 45 children who were recruited, five were dropped because English was a recent second language for them. An additional child was dropped because of prolonged absence from school. The remaining 39 children were individually given the Peabody Picture Vocabulary Test (PPVT) (Dunn, 1959) and the reading, spelling, and arithmetic subtests of the Wide Range Achievement Test (WRAT) (Jastak & Jastak, 1965). An additional six children were then dropped from the study because their PPVT IQ was below 90. None of the remaining children had any noticeable articulatory problems.

On the basis of their scores on the reading subtest of the WRAT, the 33 children were divided by reading score into three nonoverlapping groups. The 10 children (5 females, 5 males) with a reading grade level of 3.9 or below (range: 2.5 to 3.9) were designated the "poor" readers. Although the WRAT indicated that some of these children were reading at grade level, all of them were achieving below local norms, and all of them lagged substantially behind their peers. The 12 children (4 females, 8 males) with a grade level of 4.1 to 5.1 were assigned to the "average" reader group. Finally, the remaining 11 children (8 females, 3 males) with a reading level above 5.1 (range: 5.5 to 6.8) were designated the "good" readers. The mean age and test scores for each reading group are summarized in Table 1. From the table, it can be seen that the reading groups differed not only in reading level, $F(2,30) = 98.6$, $p < .001$, but also in spelling ability, $F(2,30) = 33.8$, $p < .001$. All three groups obtained grade-level scores in arithmetic. Differences between the groups, though small, were consistent enough to reach significance, $F(2,30) = 4.6$, $p < .02$. There were no significant differences in age, $F < 1$, or in IQ, $F(2,30) = 3.2$, $p > .05$.

Table 1

Experiment 1: Mean Scores of the Children as a Function of Reading Ability

	Reading Ability		
	Good	Average	Poor
n	11	12	10
WRAT grade level			
Reading	6.3	4.7	3.1
Spelling	5.5	4.5	3.0
Arithmetic	3.6	3.2	3.0
Age (yr-mon)	8-8	8-9	8-8
PPVT			
IQ	117	107	106
Raw score	80	74	72

In addition to its use in determining IQ, the PPVT was used to assess whether there were group differences in receptive vocabulary. For this comparison, the raw score (the absolute number of drawings that were recognized, unadjusted for age) of each child was examined. It was found that the groups were not equivalent on this measure, $F(2,30) = 4.8$, $p < .02$; there was a relationship between reading ability and the number of drawings recognized on the PPVT.

Materials

Forty pictured objects were selected from among the 85 line drawings of the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1976). The BNT was standardized on a group of children ranging in age from 6 to 14. The test objects were ranked by the frequency with which naming errors occurred in the standardization group, thus giving a difficulty rank to each. The "correct name" for each object was determined by consensus of educated adults. The correlation between the ranked "difficulty" (i.e., incidence of naming errors) of the objects and the frequency of occurrence of object names[3] (Carroll, Davies, & Richman, 1971) was highly significant, $r(83) = -.35$, $p < .001$. The particular objects for this study were selected from across the entire range of the BNT. An attempt was made, within the constraints of the BNT, to include objects that are difficult to name but have short names, as well as objects with long names that are easy to name. Eighteen two-syllable names were represented, along with 12 with greater than two syllables and 10 consisting of one syllable. The items chosen are listed in Appendix A along with BNT difficulty rank, number of syllables, and frequency per million words (Carroll et al., 1971).

For the naming test, the 40 pictured objects were photographed and mounted on 2 x 2-in. slides. For the recognition test, the 40 objects were reduced in size to approximately 3 x 4-in. The 40 reduced drawings were then divided into eight groups of five, all close in difficulty rank. To each group was added another three reduced BNT object drawings that had difficulty ranks near those of the original five objects. This procedure resulted in eight recognition sets, each consisting of eight pictured objects of similar BNT difficulty rank. The eight members of each set were mounted in random order on a sheet of 8 1/2 x 11-in. white paper.

Procedure

The children were tested individually in one 30-min session. For the naming test, the pictured objects were projected onto a plain white screen using a carousel slide projector. The children viewed the objects from a distance of about 52 in., with each object subtending a visual angle of approximately 5.5 degrees both vertically and horizontally. The onset of the visual display triggered the start of a clock, which was stopped by the child's vocal response, via a hand-held microphone and a voice-activated relay. The experimenter recorded all responses and the naming times of the correct responses. The entire naming test was recorded on audiotape.

At the beginning of the experiment, the child was instructed to name each object as quickly as possible. The objects were then presented sequentially in the order that they appear in the BNT, i.e., according to their rank difficulty. If the child's first response was incorrect, the experimenter asked for another name for the object. If the second response was also

incorrect, the experimenter tried to elicit a third attempt. If a child continued to respond inaccurately or gave no response at all, then his or her familiarity with the pictured object was assessed. To evaluate familiarity with an item, the experimenter asked the subject to describe the object's uses or where it had been seen before. The question was phrased in the way that was most appropriate for the particular object. If the child could demonstrate familiarity with an object, then he or she was tested for awareness of phonological properties of the name. To do this, the experimenter asked whether the object name was a short word like "cat," a medium-length word like "pencil," or a long word like "bicycle." Finally, a prompt was given consisting of the initial phonemes of the name, if the child had not already produced an incorrect response that included these phonemes. The prompt for "wreath," for example, was "/ri/."

The recognition test was conducted at the end of the test session. At that time, the child was shown each of the sets of recognition objects and was instructed to point to the object named by the experimenter. The experimenter then named in random order the eight objects of each set and recorded the subject's responses.

Results

Naming

An object was scored as correctly named if at any time its name was spontaneously given. Thus, the overall scoring did not reflect whether the name was produced on the first, second, or third try. Only a few objects were initially named correctly by a majority of the children. As a consequence, naming times on most of the objects were unavailable for most children and could not be subjected to statistical analysis. It was noted, however, that no tradeoff between speed of response and accuracy of response was evident; initial correct responses were generally given quickly. It was also noted that the stress on speed of response did not increase the likelihood that children would make errors that are phonetically related to the correct responses. Incorrect initial naming attempts bore as close a phonetic resemblance to the correct name as incorrect responses made on the second or third try when the stress on speed was relaxed.

Relationship between reading ability and object-naming ability. The number of objects correctly named without prompting ranged from as few as 10 of the 40 objects to as many as 30. The correlation between the number of objects a child named and his or her reading score proved to be significant, $r(31) = .46$, $p < .008$. Thus, there is a significant relationship between reading ability and object-naming ability.

The question arises, however, whether the poor readers named fewer objects than the good readers because they had smaller vocabularies including fewer of the object names. To examine this possibility, the results of the recognition and object familiarity testing were used to adjust each child's naming score. For the purpose of computing the adjusted score, pictured objects that were judged unfamiliar or were not recognized from their spoken names were eliminated from consideration on an individual basis. Moreover, the final five items (scroll, noose, tongs, sphinx, visor) were eliminated, because these were consistently found to be either unfamiliar or not recognizable by name. Of the remaining objects, the proportion correctly

named ranged from .34 to .94. The relationship between the proportion of objects named and the child's reading score yielded a significant correlation, $r(31) = .48$, $p < .005$. This correlation is of about the same magnitude as the value obtained when the naming score was not adjusted for object familiarity or object-name familiarity. Thus, the variation in object-naming ability with reading level could not be explained as an artifact of differences in vocabulary size; it was also obtained when the analysis was limited to familiar objects that were immediately recognized when named by the experimenter.

The effect of difficulty rank and the length of object names on naming success. Other factors in addition to reading ability may have a relationship to naming success, viz., an object's difficulty rank and the length of its name. In examining these possibilities, only the objects that were both familiar and recognizable by name were considered for each child. Since it was necessary to eliminate the final five objects, and since objects with two-syllable names were overrepresented in the stimulus set, the data were reorganized into two difficulty levels, each containing short and long names, thus comprising four groups in all. The "easy" level consisted of the first 18 objects (from "toothbrush" to "harmonica" in Appendix A). The "hard" level was composed of the next 17 objects (from "igloo" to "pyramid"). Within each difficulty level, the objects were divided by the number of syllables in their names; objects with one- or two-syllable names were said to have "short" names, whereas objects with three- or four-syllable names were said to have "long" names. For each child, the percentage of objects correctly named in each of the four groups was calculated. The mean percentages for each group are shown in Figure 1 as a function of reading ability.

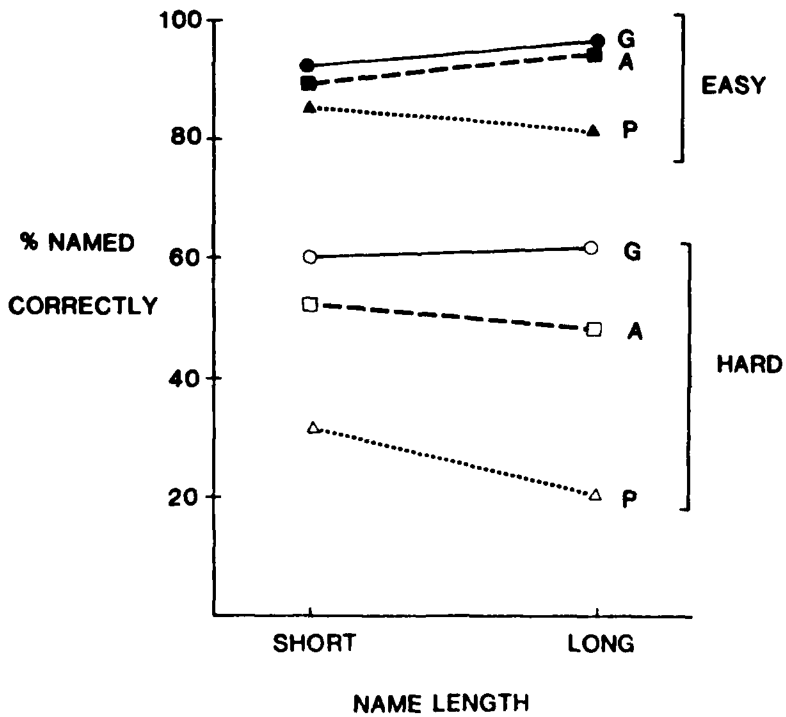


Figure 1. Experiment 1: Mean percentage of objects named correctly as a function of reading group (G = good, A = average, P = poor), difficulty level (Easy, Hard), and name length.

It is clear from inspection of the figure that naming performance varied with both reading ability and difficulty level. Furthermore, naming performance varied with reading ability to a much greater extent on the hard objects than on the easy objects. Word length appeared to have had less effect on naming than did difficulty level. For all the children, the objects with long names could be named about as well as those with short names. For the poor readers, however, there was a drop in performance on objects with long names, particularly in the hard group.

To test these observations, an analysis of variance was conducted with one between-groups factor (reading ability) and two within-groups factors (difficulty level and name length). The analysis revealed significant main effects of reading group, $F(2,30) = 7.0$, $p = .004$, and difficulty level, $F(1,30) = 300.6$, $p < .001$, and a significant interaction of the two, $F(2,30) = 5.1$, $p < .02$. Furthermore, the interaction of difficulty level and name length proved significant, $F(1,30) = 6.3$, $p < .02$.[4] The interaction of name length and reading group approached significance, $F(2,30) = 2.8$, $p < .08$.

To ascertain whether the interaction between reading ability and difficulty level might be explained as a function of absolute error scores, we can turn to a correlation measure, which is not affected by changes in scale or absolute magnitude (Baron & Treiman, 1980). Such an analysis can be meaningfully applied to the data, since reliability was comparable for the two difficulty levels. Split-half reliability adjusted by the Spearman-Brown correction was .83 for the easy objects and .86 for the hard objects. Proceeding with the analysis of the interaction, the correlation between the children's reading scores and mean performance on the difficult objects was found to be greater than that between reading scores and mean performance on the easy objects. The two correlations are, respectively, $r(31) = .50$, $p < .003$, and $r(31) = .26$, $p > .05$. (The relationship between performance on the two tasks is $r(31) = .62$, $p < .001$.) Using a formula for comparing dependent correlations (Cohen & Cohen, 1975), the two significantly differed in a one-tailed test, $t(30) = 1.8$, $p < .05$. Thus, the interaction between reading ability and difficulty level cannot be attributed to a scaling problem.

The data were also analyzed with respect to the word frequency of the object names instead of the objects' BNT difficulty ranks. Although the difficulty ranks and the word frequencies significantly correlate, the relationship is not a perfect one. On the one hand, the difficulty ranks may, perhaps, better reflect the frequency of occurrence of the object names in spoken language than the word count frequencies, which were compiled from written material. On the other hand, it is likely that the difficulty ranks are contaminated by extraneous factors, such as the ease of articulation of the object names and the quality of the object drawings themselves. Thus, the analysis based on word frequency may be as meaningful as the previous one that used difficulty rank as a factor. This analysis revealed main effects of reading ability, $F(2,30) = 8.6$, $p = .002$, frequency, $F(1,30) = 147.5$, $p < .001$, and name length, $F(1,30) = 26.2$, $p < .001$. Moreover in this analysis, the interaction between reading ability and name length was significant, $F(2,30) = 8.0$, $p = .002$; the poor readers experienced increasing difficulty labeling objects with longer names.

Error Analysis

Phonetic relationships between the errors and the target items. When an error in naming occurred, the frequency of the incorrect response word was greater than that of the target word 77% of the time. Moreover, many of the errors also bore an obvious phonetic relationship to the correct word. Examples are shown in Table 2 under the heading Word errors. In these examples, the error often shares with the target word the same stress pattern, the same number of syllables, and several phonemes. Although nonword responses were infrequent, they usually bore a strong phonetic resemblance to the target words, as is apparent in the examples given in Table 2.

Table 2

Experiment 1: Examples of Errors that Bear a Strong Phonetic Resemblance to the Target Names

Target	Word errors	Nonword errors
volcano	tornado	/blou'keiən/
		/bal'keinou/
globe	bulb	/glouv/
		/gʌlb/
harmonica	thermometer	/hə'manəkɔrn/
		/man'kanə/
stethoscope	microscope	/'sɪspəskoup/
	telescope	/'tɛθəskoup/
rhinoceros		/'rainəsɔrəs/
		/rai'nasɪs/
		/'rainəs/
		/də'ranəsɔrəs/
dominoes		/'danəmouz/
		/də'manəmouz/

The effect of reading ability. It was important to quantify the degree of phonetic relationship between the errors and the correct names in order to make comparisons across reading groups. To do so, two separate analyses were done using the initial responses on those trials on which the objects were named incorrectly. The outcome of these analyses showed no significant differences between the groups. First, the agreement between the number of syllables in the incorrect response and the number of syllables in the target name was determined. Of the 170 responses, syllable agreement occurred for 48% (effect of reading group, $F < 1$). In the second analysis, it was found that 25% of the errors, on average, had the same initial phoneme as the target names. Again, even though the poor readers had produced significantly more errors than the better readers, there was no effect of reader group, $F(2,29) = 1.5$, $p = .23$.

Familiarity with Pictured Objects

An assessment was made of the children's familiarity with the objects that were named incorrectly, as described in the Procedure. Of the 40 objects, only 2.7 were unfamiliar on average. There were no differences in object familiarity across reading groups, $F < 1$.

Tacit Knowledge of Names that Were Not Produced

If an object was incorrectly named but was nevertheless familiar, the child was asked to choose a comparison word that matched the approximate length of the correct name, as described in the Procedure. If, for example, the child had selected the word "cat," then his or her choice was a one-syllable word; if "pencil," a two-syllable word, and if "bicycle," then a three-syllable word. Agreement between the number of syllables in the target object names and the number contained in the children's choices was in this way determined. (Since a four-syllable comparison word was not available, four-syllable names were grouped with three-syllable names for this analysis.) It was found that agreement on the number of syllables tended to be low when the objects had one-syllable names. Apparently, there was a bias to choose the two-syllable item. Nevertheless, children correctly indicated the number of syllables for target items they could not produce on 63% of the trials. This percentage did not vary with reading group, $F < 1$. Thus, the children's tacit knowledge of names that were not produced was in that respect equivalent.

Effects of Prompting

In cases of failure to name an object, the child was subsequently given a phonetic prompt if he or she had passed the test of object familiarity. The prompt led to a correct response 34% of the time, on average, and the reading groups did not differ on this measure, $F(2,30) = 1.4$. One may then assess how closely related phonetically the incorrect responses were to the target names. When a prompt was ineffective, the child often failed to respond at all. When prompting elicited a response, it was often a nonword that bore a clear phonetic relationship to the target names. For example, in response to the prompt "/stɛ/" for "stethoscope," the following errors were produced: /'stɛfəkoup/, /'stɛləkoup/, /'stɛləskoup/, /'stɛpəskoup/, /'stɛsəfoun/, /'stɛlIkəl/. Again, it was desirable to quantify the phonetic relationship between the errors and the correct words in order to compare the reading groups. The incorrect responses always shared the initial phonemes with the target names because these were given as the prompt. It was determined that 66% of the cases also had the same number of syllables as the target words. Syllable agreement did not vary with reading ability.

Recognition of Pictured Objects from Spoken Names

Few errors were made in recognition of the pictured objects during the post-test when their names were spoken by the experimenter. Moreover, the percentage of correct recognitions varied only slightly with reading level; 86% of the objects were recognized by the poor readers, 88% by the average readers, and 90% by the good readers. These differences did not reach statistical significance in an analysis of variance, $F(2,30) = 2.8$, $p < .08$. In a more fine-grained analysis, however, the correlation between the children's reading scores and the number of objects recognized was

significant, $r(31) = .46$, $p < .008$. Thus, these results are consistent with the variation in receptive vocabulary with reading level found earlier using the PPVT raw scores.

Discussion

The purpose of this experiment was to examine beginning readers' naming performance in order to confirm the presence of naming deficits in poor readers and to determine whether phonological deficiencies can account for the deficits. The results showed that there is indeed a relationship between reading ability and object naming in these children. The poor readers named significantly fewer objects than either the average or the good readers. Moreover, the difference remains when the children's naming scores were adjusted by eliminating objects that were unfamiliar or those whose names were unfamiliar. Therefore, we can be confident that the relationship between reading level and naming cannot be attributed to differences either in the children's familiarity with objects or in the relative size of their recognition vocabularies.

It is plausible that the better readers had previously been exposed to many of the object names in print. Possibly, having read the object names repeatedly, the good readers' representations of the names could have been more elaborate than those of the poor readers, thus allowing the good readers to name more objects correctly. It is possible, therefore, that reading experience resulted in an improvement in the ability of the better readers to name objects. In practice, the effect of reading experience on object-naming ability is impossible to estimate. On the one hand, the better readers knew more of the words on the PPVT than the poor readers. On the other hand, the "true" effect due to reading experience might well have been slight since the children had been reading for only a short time (about a year and a half) prior to their participation in this experiment.

It is now appropriate to consider whether the naming deficits of poor readers can reasonably be attributed at least in part to deficiencies in phonological processing. First, we should note that an interaction of difficulty level and reading group was obtained, which is in keeping with the findings of Denckla and Rudel (1976). We turn to consider the interpretation of this interaction. On one account, the poor readers may have had difficulty locating phonological representations, especially those of uncommon words, possibly due to inadequate perceptual or semantic interpretation of the objects themselves. On another account of the interaction, uncommon names, having been heard less frequently, may be represented incompletely or their representations may be processed ineffectively by all the children. The representation and processing of these names may be especially deficient in the poor readers who, because of their hypothesized phonological deficiencies, may require more experience to establish usable phonological representations (and to process these representations for output), accounting for their inferior performance on naming objects with uncommon names.

If phonological deficiencies do underlie naming deficits, then other results would follow. An expected consequence of phonological deficiencies might be special difficulty naming objects with long names, since the longer the name the more phonological information that must be represented and processed. In this regard, the interaction between name length and reading group is of interest. It approached significance when it was analyzed in

conjunction with difficulty level, as assessed by the BNT ranks, and it attained significance when frequency in print of the object names was used as a factor instead of BNT rank. An increase in error rate on longer names cannot readily be accounted for by a general perceptual or semantic deficiency leading to difficulty locating phonological representations. Such a problem should be insensitive to the length of the objects' names. Furthermore, the poor readers' difficulty with long names cannot be accounted for by supposing that they have an articulatory problem that hinders their production of long names. The poor readers were able to label correctly about half the objects that had long names, and their erroneous responses were sometimes long words. In view of its importance in explaining the naming deficits of poor readers, the relationship betweeen name length and reading ability merits further investigation.

The results of the error analysis indicated that the incorrect responses of all the children, irrespective of their reading level, were equivalent in degree of phonetic relationship (as judged by the initial phoneme and word length) to the correct object names. Moreover, all the children, by producing incorrect responses that were phonetically related to the correct names, demonstrated that they could locate the correct phonological representations and that some of the phonological information was brought to bear in articulating their responses. When errors in naming occurred, we may suppose that the representations were not sufficiently detailed or not effectively processed. The results of the error analyses reveal no problems peculiar to the poor readers, but their higher error rate is consistent with the many sources of data that implicate phonological immaturity and deficient processing in this group.

Further evidence that implicates phonological deficiencies resulted from tests of the children's awareness of object names that were not correctly produced. Awareness of the length of the object names was above chance and did not vary with reading level. This is consistent with the results of Wolf (1979), who employed a similar procedure. This result should be interpreted cautiously, however, since the children usually did not offer a response on every trial of this task. If the children had been required to respond on every trial in which they failed to name the object correctly, they might have registered a lower level of accuracy and performances might have varied with reading ability. Nonetheless, the present findings are compatible with the idea that all the children could locate the appropriate phonological representations and that word length was specified in the representations. However, it might be supposed that full segmental information was not represented completely enough to enable the children to carry out the processing necessary to produce the name.

Finally, there were no differences across reading groups in sensitivity to phonetic prompts. The likely effects of prompting are complex. It is possible that the prompt, by providing speech cues, aided all the children in finding the correct phonological representation. On the other hand, it may be that the prompt provided confirmatory evidence that the children had found the correct representation. Following that, they may have been less reluctant to use the specified information. In either case, the high incidence of nonword responses after prompting indicates that many of the phonological representations contained partially deficient segmental information, although word length was relatively well represented.

Qualitative differences between the groups did not emerge from the error analysis or in the response to various probes for tacit knowledge of the properties of misnamed items. Apparently, when the good and the average readers failed to name pictures, their failures were similarly determined. The reading groups differed, however, in how often they were able to use their representations of names to produce the standard labels for the stimulus objects. Thus, the results of this experiment provide support for the hypothesis that the poor readers had difficulties naming objects because of underlying deficiencies in representing phonological information and in generating responses from the phonological representations.

Experiment 2

In Experiment 1, the evidence for phonological deficiencies was provided by using an object-naming task. Object naming, like speaking spontaneously, requires that phonological representations be used to guide the overt production of the target word. Use of phonological representations in this way is obviously a well-practiced routine, and humans are specially equipped biologically to carry it out (Lenneberg, 1967). The use of phonological representations in other ways, however, may require linguistic abilities different from those necessary for speaking. More specifically, making metalinguistic decisions based on the characteristics of words requires an explicit awareness of the phonological composition of those words, an awareness that is not necessary for normal speaking, but may be necessary for effectively learning to read language that is written by an alphabet (Liberman et al., 1977). Moreover, if the metalinguistic decisions are to be made on the names of objects, then the ability of subjects to use phonological representations that are stored in long-term memory can be assessed. The present experiment explores the possibility that poor readers would prove deficient at using phonological representations to perform metalinguistic decisions even on words whose representations are completely specified in long-term memory.

The requirement that metalinguistic decisions be based on stored phonological representations may make for greater difficulty than the same decisions based on words presented auditorily. In fact, it is possible that certain metalinguistic tasks could be done easily by poor readers on spoken words, but only with great difficulty when they are required to generate the necessary phonological information without the acoustic cues provided by speech. Judging the length of a pair of words and deciding whether two words rhyme are metalinguistic tasks that are within the capability of young children when words are presented auditorily. In this connection, it has been found that 90% of the children in a first-grade class could indicate correctly the number of syllables in words presented auditorily (Liberman et al., 1974). The number of syllables in a word is, of course, a good measure of its relative length. Thus, the information necessary to judge the length of a spoken word is available to children before they reach school age. Moreover, rhyme is a phonological relationship that is easy for young children to identify in spoken words (Lenel & Cantor, 1981). Thus, the two questions being asked in themselves are not likely to be beyond the abilities of the subjects.

Even though young children are able to make rhyme decisions and length decisions on spoken words, the same decisions may be difficult when they have to be based on representations that must be accessed through some other medium

than that of speech, as, for example, the medium of pictures. Decisions based on object names require that the necessary phonological characteristics of the names be adequately represented in long-term memory. Experiment 1 suggested that poor readers may be deficient at representing the full segmental structure of words, although they may be able to represent adequately their gross characteristics, including approximate length. Since rhyme decisions based on object names apparently require that the full segmental structure be represented, it would come as no surprise if poor readers were deficient in making these decisions. In contrast, poor readers would not necessarily be deficient in making decisions based on word length provided that they could become explicitly aware of this attribute. The issue of the children's awareness of the length of object names not produced was examined incompletely in Experiment 1 and did not produce a clearcut result.

Thus, two difficulties would lead to deficient performance on certain metalinguistic tasks: a difficulty in representing the pertinent attributes of words and a lack of awareness of those attributes, which must become explicitly known in order to carry out the tasks. To examine whether the second possibility is indeed a genuine problem, we must first ascertain that the necessary information about a word is represented completely. Proof that a word is well-represented phonologically is demonstrated by the ability to generate the word acceptably. Accordingly, in this experiment the children were asked to perform metalinguistic tasks requiring access to the names of objects. It was later investigated to what extent the names were represented completely by testing for the ability to name the objects aloud. Following that, consideration was restricted to those item presentations for which it could thus be shown that the names were adequately represented. If the performance of poor readers was shown to be inferior to that of good readers even on these presentations, then evidence will have been adduced that poor readers lack explicit awareness of certain phonological properties of words they know.

Method

Subjects

The subjects were the children who participated in Experiment 1. Two children (a boy reading at a 5.5 grade level and a girl with a 6.8 reading level) were dropped from the study due to prolonged absence from school. Despite the loss of these two subjects, the test scores of the present group of good readers were close to those of the group described earlier (see Table 1).

Materials

For the rhyme condition, 15 pairs of line drawings of objects with rhyming names and 15 pairs with nonrhyming names were prepared. The names in each pair were monosyllabic words matched in frequency of occurrence[5] (Carroll et al., 1971). In addition, the mean frequency of each rhyming pair of names approximated the mean frequency of one of the nonrhyming pairs. (The names of the objects are listed in the left side of Appendix B. The first pair in each column was used for practice.)

For the length condition, 15 pairs of line drawings of objects with monosyllabic names were prepared. As a control, an additional 15 pairs of pictured objects with names of different length were also prepared. For the latter, one object in each pair had a monosyllabic name and the second object had a polysyllabic name, usually comprising three syllables. As in the rhyme condition, the names of the two objects in each pair were matched in frequency of occurrence. Further, the mean frequencies of the same-length pairs were matched to those of the different-length pairs. Moreover, each pair in the length condition was matched in frequency to a pair in the rhyme condition. (The names of the objects used in the length condition are listed in Appendix B. Again, the first pair in each column was used for practice.)

The two pictured objects designated for each test trial were separated by a vertical line, photographed, and mounted on 2 x 2-in. slides. For the different-length series, the object with the long name appeared on the left on half the slides and on the right for the other slides. The order of the slides in the rhyme condition was random with the constraint that no more than three successive trials be either rhyme or nonrhyme trials. The same ordering was used for the slides in the length condition.

<u>Procedure</u>

The children were tested individually on both the rhyme and length conditions in a single 30-min session. The order of conditions was counterbalanced so that half the children in each reading group received the rhyme condition first and the length condition second. The order of conditions was reversed for the remaining children.

In each condition, the pictured objects were projected onto a plain white screen using a carousel slide projector. The onset of the visual display triggered the start of a clock, which was stopped when the child pressed one of two telegraph keys. The children viewed the pictured objects from a distance of approximately 52 in., and each object subtended a visual angle of approximately 4.4 degrees both vertically and horizontally.

For the rhyme condition, the experimenter first ascertained that the child could distinguish spoken rhyming words and nonrhyming words. The experimenter spoke pairs of words and asked the child if they rhymed. Following that, the child was told that two pictured objects would appear simultaneously on the screen and that the task was to indicate quickly whether the objects had rhyming names. Each subject responded by pressing either the key labeled "YES" or the key labeled "NO." As a reminder of the task, a card marked "Rhyme?" was placed between the keys. The child's responses on the two practice trials were reviewed to ensure that the task was understood.

For the length condition, it was first ascertained that the child could distinguish spoken monosyllabic and polysyllabic words by indicating whether words spoken by the experimenter were "long" or "short." Then pairs of words were given and the child had to indicate whether or not both words were short. Following this pretest, the subjects were asked to make length judgments on pairs of pictured items. The task was to indicate as quickly as possible whether the names of two pictured objects presented simultaneously were both short (i.e., monosyllabic). The child again responded by pressing one of two keys, one labeled "YES" and the other "NO." As a reminder of the task, a card marked "Both short?" was placed between the keys. As in the rhyme condition,

two practice trials preceded the test trials and the subject's responses were reviewed.

Following the testing on both conditions, the children were again shown each test slide. This time they were asked to name the objects aloud.

Results

For each task, the mean percentage of correct responses and the mean response times on correct trials were calculated. These calculations were made separately for the trials on which the correct answer was "no" (the so-called "no" trials) and for the trials on which the correct answer was "yes" (the "yes" trials). Because of the error rate, it was not practical to subject the response times to statistical analysis. The mean percentages of correct responses are shown in Figure 2 as a function of reading ability and task. When one examines the data from the "no" trials alone (left graph), it can be seen that overall performance on the rhyme task was very accurate; indeed, all the children performed near the ceiling level. In contrast, on the length task, performance varied markedly with reading ability. An analysis of variance with one between-groups factor (reading ability) and one within-groups factor (task) was conducted. In accordance with the above observations, main effects of reading group, $F(2,28) = 15.0$, $p < .001$, and task, $F(1,28) = 53.5$, $p < .001$, were obtained. Moreover, there was a significant interaction between reading ability and task, $F(2,28) = 7.6$, $p = .003$.

The mean percentages of correct responses on the "yes" trials are also displayed in Figure 2 (right graph). Compared with the corresponding percentages on the "no" trials, these values were generally lower. Neither the length task nor the rhyme task is near the ceiling level. It is apparent from the table that overall accuracy varied as a function of reading ability; the poor readers were correct on 64% of the trials, the average readers on 77%, and the good readers on 79%. Performance on the two tasks was comparable in overall accuracy with 74% correct on each, but varied with reading ability, particularly on the length task.

To evaluate these differences statistically, an analysis of variance analogous to that for the "no" trials was conducted. The analysis revealed a main effect of reading ability, $F(1,28) = 7.3$, $p = .003$. The interaction between reading group and task also proved significant, $F(2,28) = 6.3$, $p = .006$. The poor readers again had special difficulty on the length task even though all the object names on the "yes" trials were monosyllabic words.

Two possibilities come to mind as explanations of the inferior performance of the poor readers on the length task. Obviously, if their representation of word length information were inadequate, then the poor readers would fail to make correct length decisions. Even with adequate representations, however, difficulties could arise if the poor readers were unable readily to become aware of the word length specified by the representations. To investigate this one must first have ensured that any given subject's representation of word length is accurate. To that end, each child's task performance was assessed using only those trials on which both pictured objects had been later named correctly. For items that meet this criterion, the object names must have been represented entirely. Thus, if the poor readers prove to have difficulty making length decisions on these object

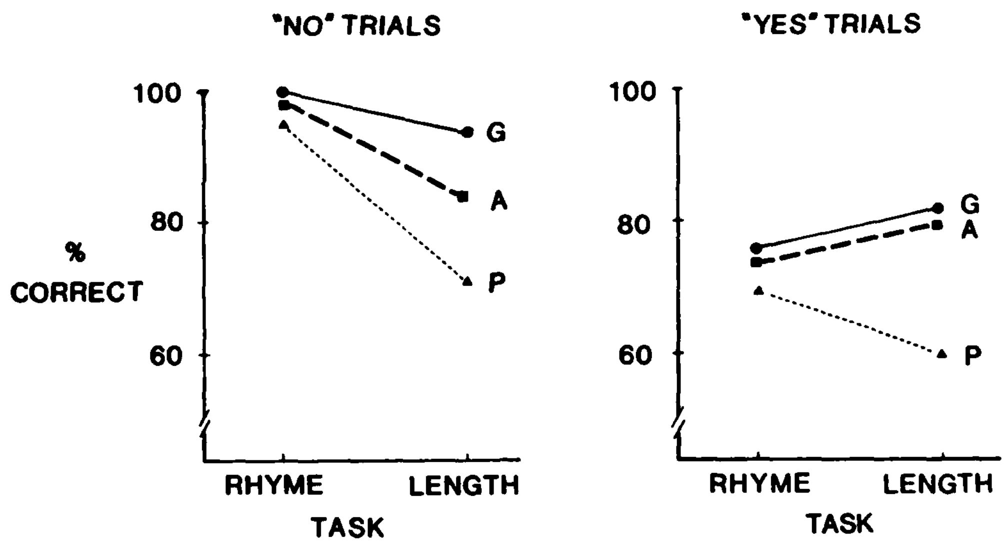


Figure 2. Experiment 2: Mean percent correct as a function of reading ability (G = good, A = average, P = poor) and task.

names, their failure must indicate a lack of awareness of the length of the object names specified by these representations.

Considering only those trials on which the objects could be named, the mean percentages of correct decisions are shown in Figure 3. On both the "yes" and the "no" trials, it can be seen that performance was very accurate for all children on the rhyme task, but that it varied with reading ability on the length task. The effects of reading ability and task and their interaction were computed and are given in that order: for the "yes" trials, $F(2,28) = 3.9$, $p = .032$, $F(1,28) = 18.6$, $p = .001$, and $F(2,28) = 7.1$, $p = .004$; for the "no" trials, $F(2,28) = 17.5$, $p < .001$, $F(1,28) = 51.0$, $p < .001$, and $F(2,28) = 10.0$, $p < .001$. Possibly, the interaction effects in these analyses were inflated, since rhyme performance approached ceiling levels. Nevertheless, it is clear that performance on the length task effectively distinguished the reading groups. Analyses of variance with one factor (reading ability) computed on only the length task data were highly significant; for the "yes" trials, $F(2,28) = 8.5$, $p < .002$; for the "no" trials, $F(2,28) = 14.6$, $p < .001$.

Thus it is found that even when the children demonstrated that they could name both objects on the length task, the poor readers nonetheless failed more often than the good readers to make accurate decisions. Therefore, one may

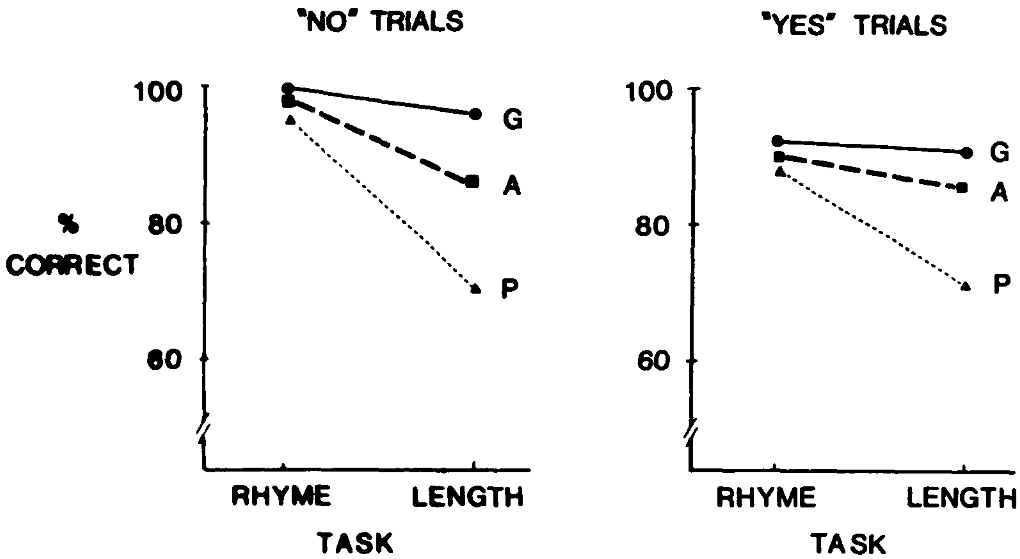


Figure 3. Experiment 2: Mean percent correct as a function of reading ability (G = good, A = average, P = poor) and task when the objects were nameable.

suppose that the poor readers found it particularly difficult to make explicit the word length information specified in a phonological representation.

Discussion

The purpose of this experiment was to explore the possibility that poor readers are deficient in using their phonological representations to guide performance on two metalinguistic tasks: a rhyme task, which required them to decide whether two objects have rhyming names, and a length task, which required them to decide whether two objects both have short names. The results indicated that the relationship between performance on the rhyme task and reading ability was small. There was, in contrast, a strong relationship between performance on the length task and reading ability. Considering only those trials on which objects were successfully named, performance on the length task improved for all the subjects, but the poor readers' performance remained significantly inferior to that of the better readers. Therefore, it can be said that the poor readers have a genuine difficulty in making length decisions even on words that are fully represented in long-term memory.

The results of Experiment 2 raise several issues. To begin with, examining only the trials on which both objects could be named, we see that complete representation of the object names provided a firm basis for making

accurate rhyme decisions for all the children. The high level of performance indicates as well that, by the third grade, rhyme is a very salient characteristic of words. Children, of course, are acquainted with the existence of rhyming words, since these occur often in children's verse and song. The children's ability to make rhyme decisions on object names that vary in completeness of representation can be examined by considering all the trials (not just those that presented objects that could be named correctly). On the "no" trials of the rhyme task, all the children performed at high levels of accuracy, whereas on the "yes" trials, performance was at lower levels. This finding supports the view that rather complete representation is necessary for subjects to recognize that object names rhyme, but that incomplete representation provides an adequate basis for deciding that they do not. Apparently, incomplete representation of object names existed even for the good readers sufficiently to lower response accuracy on the "yes" trials.

Although the poor readers performed as well on the rhyme task as the better readers, they were unable to become explicitly aware of the length of words that were represented in memory. This finding is ostensibly discrepant from the result of an awareness test that was conducted in Experiment 1. In that experiment, when an object was familiar but could not be named, the child was asked to decide whether the object name was a short word like "cat," a medium-length word like "pencil," or a long word like "bicycle." It was found that reading groups were not differentiated on this task. This result, however, must not be overinterpreted, since the children often failed to respond on these occasions. Caution in interpreting the earlier finding is reinforced by the results of the present experiment.

Additionally, it may be that the length task in the present experiment was particularly taxing for the subjects. It required them to use their internal representations to judge the lengths of each pair of test words, whereas the task in Experiment 1 required only that the subject assess the length of a single word from lexically represented information. A further procedural difference that could have contributed to the difference in outcome of the two experiments was the provision of a spoken comparison word in Experiment 1. In that experiment, the children were asked to match the length of an object name with one of three words spoken by the experimenter. By being provided with explicit reference words, the children were given benchmarks that could have aided them in their length decisions. In the present experiment, a comparison was required, but no concrete standards were provided.

The discrepancy between the poor readers' use of length information in Experiment 1 compared with Experiment 2 may be viewed as an important indication of one source of difficulty among the poor readers. Thus far, the term "phonological deficiencies" has been used to encompass a deficiency in representing phonological information completely and a deficiency in the processing applied to the representations. Since representations and processes applied to them are interdependent (Anderson, 1978; Palmer, 1978), it can be difficult to distinguish between deficiencies in the two components. In Experiment 1, moreover, it was reasonable to consider the two deficiencies together since they would have the same effect on object-naming performance. Both deficiencies would become manifest after a particular phonological representation had been located. However, a comparison of the poor readers' use of length information in Experiment 1 with that in Experiment 2 indicates that, whatever difficulty the poor readers may have had in representing

phonological information fully, they also had a problem using adequately represented information to perform particular metalinguistic tasks. In Experiment 1, the poor readers, like the better readers, were able to use stored phonological information to produce naming responses that, although incorrect, matched the target names in length. In Experiment 2, however, the poor readers had difficulty processing the stored phonological information in order to respond accurately on the metalinguistic length task.

One may ask why the metalinguistic length decisions of Experiment 2 so effectively differentiated the reading groups. The question is the more pertinent in view of the results of Liberman et al. (1974) that showed that poor readers can demonstrate their awareness of the length of spoken words by indicating the number of syllables in each. That study showed, moreover, that a matched group of children could not do the more difficult task of indicating the number of phonemes in spoken words. Those findings are among several indications that poor readers lack explicit knowledge of the phonemic units of spoken words (Alegria, Pignot, & Morais, 1982; Treiman & Baron, 1981). In the present experiment, the length task could have been done successfully using either syllabic or phonemic information. Nevertheless, the poor readers could not judge the lengths of words when they had to depend solely on the phonological representations stored in long-term memory in order to generate the necessary information. It is plausible that the poor readers failed on this task because they lacked explicit awareness of the units of their phonological representations, which correspond to the units of spoken words. Thus, although a variety of tasks (naming, reading, metalinguistic judgments) may rely on the same long-term store of phonological information, these tasks may make quite unequal demands on the processors that draw upon that stored knowledge. In keeping with the results of Liberman et al. (1974), the present study offers support for the hypothesis that poor readers generally lack an understanding of the relationship between the units of spoken words and the units of the phonological representations that underlie them. The results also support the notion (Mattingly, 1984) that a major aspect of linguistic awareness differentiating good and poor readers pertains to knowledge of mental representations.

It is to be expected that reading experience would serve to increase sensitivity to word length. There is, after all, a fairly direct relationship between the spoken length of a word and the number of letters in the orthographic form of the word. Thus reading experience could increase awareness of word length by providing a redundant cue, thereby facilitating word length judgments. Moreover, the better readers may well have seen some of the object names in print; they could have been assisted in their decisions by being able to compare the orthographic forms of the object names. The poor readers would be less able to bring this knowledge to bear on the task. In fact, it is conceivable that if the poor readers found word length decisions unduly difficult, they may have adopted an alternative strategy that was counterproductive. One possibility is that they based their length decisions on the actual sizes of the objects that were pictured rather than on the names of the objects. This possibility can, of course, be tested in the future.

General Discussion

The purpose of this two-part study was to examine how underlying phonological deficiencies could affect object naming, metalinguistic decisions, and reading. The first experiment confirmed the existence of

naming deficits in poor readers and found that their difficulties in naming are not merely a reflection of individual differences in vocabulary size. It also established a possible role of phonological deficiencies in accounting for the naming deficits. On the metalinguistic tasks of the second experiment, the poor readers were inferior to better readers in the ability to judge the relative lengths of the names of objects, even in those instances in which the children were later able to name the objects aloud. Therefore, the poor readers may lack an awareness of the word lengths specified by their internal phonological representations. The same deficiencies in the phonological domain, then, are implicated in both the object-naming deficits of poor readers and their reading deficits.

Some investigators, notably Denckla and Rudel (1976) and Wolf (1981), have compared the naming deficits of poor readers with what is known about the deficits of aphasics. From the standpoint of the present findings, one may ask specifically to what extent the naming deficits of aphasics, like those of poor readers, can be assigned to phonological deficiencies rather than to deficiency of another ability underlying the naming process. There is evidence that the problem of some aphasics occurs in attempting to locate the correct phonological representations (Mills, Knox, Juola, & Salmon, 1979; Schuell, Jenkins, & Jiménez-Pabón, 1964; Wiegel-Crump & Koenigsknecht, 1973), and this problem could be due, in principle, either to a semantic or a perceptual deficiency. However, in some cases of aphasia, as in children who are poor readers, phonological deficiencies have been implicated as a probable cause of naming failure. For example, it has been supposed (e.g., Luria, 1966) that fluent aphasics with superior temporal damage make errors on object-naming tasks partly because disintegration of phonetic analyzers leads eventually to deterioration of phonological representations. There is, in any case, evidence that aphasics, like the poor readers in the present study, often have knowledge of object names that cannot be spontaneously produced (Barton, 1971; Goodglass, Kaplan, Weintraub, & Ackerman, 1976).

Recently, a particularly compelling case of deficient phonological processing in an aphasic patient was studied in depth by Caramazza, Berndt, and Basili (1983). This individual appeared to have a normal ability to process stimuli visually and semantically, but was apparently incapable of completing any task that required phonological processing. For example, when asked to select objects with rhyming names, he performed at chance. Although this patient's phonological deficiencies were far more serious than those of the poor readers studied here, the similarities merit further comparative study.

It was suggested in the introduction that semantic errors can occur because the phonological representations of the target words are incomplete or because they cannot be processed effectively. Conceivably, many of the semantic errors that are so frequent in cases of aphasia may be due to similar phonological deficiencies. Indeed, explanations along these lines have occasionally been given in the research literature on aphasia. For example, Luria (1966) has suggested that some aphasics substitute semantically-related words on object-naming tasks because of phonological problems. Moreover, others (Baker, Blumstein, & Goodglass, 1981) have proposed that semantic errors may increase in frequency as the phonological processing required of aphasic subjects becomes more taxing. It has also been suggested that some individuals with acquired dyslexia may make semantic reading errors as a result of phonological problems occurring after the correct lexical

representation has been located (see Shallice & Warrington, 1980, for a review). The caveat that was applied to the interpretation of misnaming by children with reading disability could apply also to the interpretation of the errors made in acquired anomia: one must be wary of assuming that semantic errors imply a semantic deficiency.

We have seen how phonological deficiencies in processing information stored in long-term memory can lead to errors in naming. Poor readers also have short-term memory problems that are specific to the retention of phonetic material (Liberman et al., 1977; Shankweiler et al., 1979). It was suggested (Shankweiler et al., 1979) that this phonetic memory problem could underlie other problems of poor readers that depend on the short-term retention of words, such as their difficulty remembering item order (Katz et al., 1981) and comprehending sentences (Mann, Shankweiler, & Smith, 1984). In the present study, a parallel case was made that poor readers often fail on tasks requiring knowledge of words stored in long-term memory because of underlying deficiencies in phonological abilities. The deficiencies became manifest in the two tasks of the present study that used pictured objects to elicit stored linguistic representations and corresponding spoken words.

References

Alegria, J., Pignot, E., & Morais, J. (1982). Phonetic analysis of speech and memory codes in beginning readers. Memory & Cognition, 10, 451-456.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. Psychological Review, 85, 249-277.

Baker, E., Blumstein, S. E., & Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. Neuropsychologia, 19, 1-15.

Baron, J., & Treiman, R. (1980). Some problems in the study of differences in cognitive processes. Memory & Cognition, 8, 313-321.

Barton, M. I. (1971). Recall of generic properties of words in aphasic patients. Cortex, 11, 73-82.

Benson, D. F., & Geschwind, N. (1969). The alexias. In P. J. Vinken & G. W. Bruyn (Eds.), Handbook of clinical neurology (Vol. 4). Amsterdam: North Holland.

Brady, S., Shankweiler, D., & Mann, V. (1983). Speech perception and memory coding in relation to reading ability. Journal of Experimental Child Psychology, 35, 345-367.

Caramazza, A., & Berndt, R. S. (1978). Semantic and syntactic processes in aphasia: A review of the literature. Psychological Bulletin, 85, 898-918.

Caramazza, A., Berndt, R. S., & Basili, A. G. (1983). The selective impairment of phonological processing: A case study. Brain and Language, 18, 128-174.

Carroll, J. B., Davies, P., & Richman, B. (1971). Word frequency book. New York: American Heritage.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 12, 335-359.

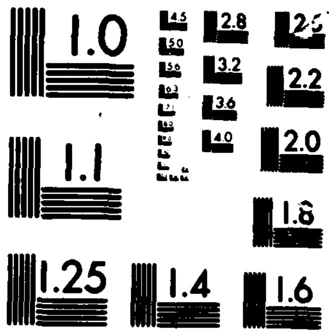
Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum.

Denckla, M. B., & Rudel, R. G. (1976). Naming of object-drawings by dyslexic and other learning disabled children. Brain and Language, 3, 1-15.

Dunn, L. M. (1959). Peabody picture vocabulary test. Circle Pines, MN: American Guidance Service.

AD-A168 819 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 3/3
AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN JAN 86 SR-84(1985) N00014-83-K-0083

UNCLASSIFIED F/G 5/7 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6

Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. Linguistic Inquiry, 8, 505-520.

Goodglass, H. (1980). Disorders of naming following brain injury. American Scientist, 68, 647-655.

Goodglass, H., Kaplan, E., Weintraub, S., & Ackerman, N. (1976). The "tip-of-the-tongue" phenomenon in aphasia. Cortex, 12, 145-153.

Jansky, J., & deHirsch, K. (1973). Preventing reading failure. New York: Harper & Row.

Jastak, J. F., & Jastak, S. R. (1965). Wide range achievement test. Wilmington, DE: Guidance Associates.

Kaplan, E., Goodglass, H., & Weintraub, S. (1976). Boston naming test.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. Journal of Experimental Child Psychology, 32, 474-484.

Labov, W. (1973). The boundaries of words and their meanings. In C.-J. N. Bailey & W. Shuy (Eds.), New ways of analyzing variation in English. Washington, DC: Georgetown University Press.

Lenel, J. C., & Cantor, J. H. (1981). Rhyme recognition and phonemic perception in young children. Journal of Psycholinguistic Research, 10, 57-67.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), Orthography, reading and dyslexia. Baltimore: University Park Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Reading and awareness of linguistic segments. Journal of Experimental Child Psychology, 18, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. (1977). Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), Toward a psychology of reading: The proceedings of the CUNY Conference. Hillsdale, NJ: Erlbaum.

Luria, A. R. (1966). Higher cortical functions in man. New York: Basic Books.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. (1980). Children's memory for sentences and word strings in relation to reading ability. Memory & Cognition, 8, 329-335.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading disability: The role of phonetic representation. Journal of Child Language, 11, 627-643.

Mattingly, I. G. (1984). Reading, linguistic awareness, and language acquisition. In J. Downing & R. Valtin (Eds.), Language awareness and learning to read. New York: Springer-Verlag.

Mattis, S., French, J. H., & Rapin, I. (1975). Dyslexia in children and young adults: Three independent neuropsychological syndromes. Developmental Medicine and Child Neurology, 17, 150-163.

Miller, G. A. (1978). Practical and lexical knowledge. In E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.

Mills, R. H., Knox, A. W., Juola, J. F., & Salmon, S. J. (1979). Cognitive loci of impairments in picture naming by aphasics. Journal of Speech and Hearing Research, 22, 73-87.

Nelson, H. E., & Warrington, E. K. (1980). An investigation of memory functions in dyslexic children. British Journal of Psychology, 71, 487-503.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum.

Schuell, H., Jenkins, J. J., & Jiménez-Pabón, E. (1964). Aphasia in adults. New York: Harper & Row.

Shallice, T., & Warrington, E. K. (1980). Single and multiple component central dyslexic syndromes. In M. Coltheart, K. Patterson, & J. C. Marshall (Eds.), Deep dyslexia. Boston: Routledge and Kegan Paul.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. Journal of Experimental Psychology: Human Learning and Memory, 5, 531-545.

Treiman, R., & Baron, J. (1981). Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), Reading research: Advances in theory and practice (Vol. 3). New York: Academic Press.

Wiegel-Crump, C., & Koenigsknecht, R. A. (1973). Tapping the lexical store of the adult aphasic: Analysis of the improvement made in word retrieval skills. Cortex, 9, 411-418.

Wolf, M. (1979). The relationship of word-finding and reading disorders in children and aphasics. Unpublished Ed.D. dissertation, Harvard University.

Wolf, M. (1981). The word-retrieval process and reading in children and aphasics. In K. Nelson (Ed.), Children's language (Vol. 3). New York: Gardner Press.

Footnotes

[1]One would also expect semantic errors to be made in instances where the correct word is not lexically represented at all. This points to the need to control for vocabulary differences in naming studies.

[2]This is a matter of concern not only in the area of childhood reading disability, but also in the aphasias of adults, where reading problems are often accompanied by naming problems (Benson & Geschwind, 1969).

[3]The frequency per million words for each name was calculated by summing the frequency of occurrence for the target word (e.g., whistle) and all syntactic variants of the name (e.g., whistles, whistled, whistling). The frequencies in the word count itself were determined by examining how often each lexical form occurred in elementary school and junior high school textbooks.

[4]It was desirable to test whether these findings can be taken to generalize to any set of objects. This was accomplished by considering the individual objects as a random effect in an analysis of variance (Clark, 1973). Since in every case but one the same effects were significant in this second analysis of variance as in the original analysis, we can be sure that the first results were not specific to any one set of objects. The analysis revealed significant main effects of reading group, difficulty level, and their interaction, respectively, $F(2,62) = 14.2$, $p < .001$, $F(1,31) = 54.7$, $p < .001$, $F(2,62) = 4.4$, $p < .02$. The interaction of difficulty level and name length was not significant in this analysis. The other results can be generalized.

[5]As in Experiment 1, the frequencies (per million words) were for the name itself and all syntactic variants of the name.

Appendix A

Experiment 1: Characteristics of Objects Selected from the *Boston Naming Test*

Object Name	Difficulty Rank	Syllables	Frequency
toothbrush	7	2	1
whistle	9	2	46
helicopter	12	4	17
mushroom	14	2	10
camel	15	2	22
wheelchair	16	2	*
octopus	18	3	3
snail	23	1	13
canoe	24	2	36
raft	25	1	18
wreath	26	1	3
plug	27	1	10
volcano	29	3	26
faucet	30	2	2
dart	32	1	5
seahorse	33	2	*
globe	34	1	35
harmonica	35	4	2
igloo	37	2	1
cactus	39	2	13
acorn	41	2	5
rhinoceros	43	4	2
dominoes	45	3	*
propeller	48	3	7
hammock	50	2	2
medal	51	2	7
unicorn	54	3	*
stethoscope	58	3	1
asparagus	60	4	1
briefcase	62	2	*
pinwheel	63	2	1
hourglass	64	2	2
nozzle	66	2	2
accordion	67	4	2
pyramid	68	3	15
scroll	69	1	2
noose	71	1	1
tongs	74	1	1
sphinx	77	1	1
visor	78	2	1

*Word frequency less than 0.5 per million.

Appendix B

Experiment 2: Stimulus Items

Rhyme	Freq	Nonrhyme	Freq	Same length	Freq	Different length	Freq
cat hat	Prac	saw house	Prac	broom comb	Prac	toothbrush tree	Prac
bear square	156	cloud heart	159	egg wheel	169	glasses bed	175
wing ring	107	dress chair	111	train church	110	apple bone	109
cake snake	65	fence desk	67	bat bow	65	spider knife	67
nail whale	60	brush pie	53	bowl pig	57	balloon bus	58
clock lock	55	pan flag	51	tank duck	56	buffalo barn	54
lamp stamp	48	chain doll	43	cage belt	44	camera pipe	46
drum thumb	31	fox pump	42	frog owl	40	banana net	30
skunk trunk	27	tent towel	29	bee seal	35	dinosaur knot	29
boot flute	22	hook rake	21	pen pear	23	typewriter ghost	21
gear spear	19	sink fork	27	drill hose	21	strawberry whip	19
clown crown	18	skis grapes	19	shark glove	20	thermometer spoon	18
kite knight	18	cane soap	21	sock mask	14	umbrella screw	13
bench wrench	12	witch dice	12	sword mop	12	butterfly harp	13
spool stool	6	braid nun	4	bride maze	5	cigarette hoe	7

ACCESS TO SPOKEN LANGUAGE AND THE ACQUISITION OF ORTHOGRAPHIC STRUCTURE: EVIDENCE FROM DEAF READERS*

Vicki L. Hanson

Abstract. Sensitivity to two types of orthographic structure was investigated: Linguistically-based orthographic regularity and summed single letter positional frequency. Deaf college students were found to make use of positional frequency information no less than hearing college students; however, the extent to which they made use of orthographic regularities in word recognition was related to their speech production skills. In one task, subjects were presented nonword letter strings for short durations, each followed by a masking stimulus and a target letter. They were asked to indicate whether or not the target letter had been present in the letter string. It was found that the accuracy of deaf subjects with good speech, like that of hearing subjects, was considerably greater for orthographically regular than irregular strings. In contrast, the accuracy of deaf subjects with poor speech was much less related to orthographic regularity. In a second task, in which subjects made judgments about how word-like various letter strings appeared, the judgments of the hearing subjects were more influenced by regularity than those of deaf subjects with poor speech. These results are discussed in terms of how expertise in speech relates to appreciation of orthographic regularity.

Introduction

It has been known for some time that hearing readers identify letters more accurately in orthographically legal nonwords (pseudowords) than in orthographically illegal nonwords (Adams, 1979; Aderman & Smith, 1971; Baron & Thurston, 1973; Gibson, Pick, Osser, & Hammond, 1962). This finding has suggested that readers of English are influenced by orthographic structure in word recognition. Orthographic structure could facilitate perception by producing constraints on letter sequences that facilitate visual processing of letter strings (e.g., Carr, Posner, Pollatsek, & Snyder, 1979; Massaro, Taylor, Venezky, Jastrzembski, & Lucas, 1980; Singer, 1980) or facilitate

*Quarterly Journal of Experimental Psychology, in press.
Acknowledgment. This research was supported by National Institutes of Health Grants NS-18010 and HD-01994. I am grateful to Rena Krakow and Nancy Fishbein for their help in stimulus preparation and subject testing, and to Tom Carr, Laurie Feldman, Carol Fowler, Alvin Liberman, John Richards, and the journal reviewers for very helpful comments on earlier versions of this manuscript. The cooperation of Gallaudet College, particularly Horace Reynolds and William McFarland, in this research project is gratefully acknowledged.

perception by allowing well-structured strings to be more readily translated into a speech representation (e.g, Spoehr & Smith, 1975).

Differences have arisen as to how to describe the nature of this structure. Descriptions have generally been divided into those based on linguistic regularity and those based on statistical redundancy (for a review, see Massaro et al., 1980). Descriptions of orthographic structure based on linguistic regularity take into account phonological and scribal constraints of English. Orthographically regular words must therefore be pronounceable and contain only legal consonant and vowel combinations: the letter string REMOND, for example, would be considered as orthographically regular and the string RMNOED would be irregular. Descriptions of orthographic structure based on statistical redundancy take into account frequency of letters or letter combinations occurring in natural text. These redundancy descriptions have taken two forms: spatial (or positional) redundancy based on counts of single letters and their positions of occurrence, and sequential redundancy based on bigram or trigram frequency counts. According to a spatial redundancy description, for example, strings high on such a measure contain letters occurring in common positions while strings low in such a measure contain letters occurring in low frequency positions.

The evidence indicates that both orthographic regularity and statistical redundancy measures describe sources of perceptual facilitation (Henderson, 1982). That is, strings that are orthographically regular are recognized more accurately than strings that are irregular (Massaro, Venezky, & Taylor, 1979; Massaro et al., 1980), and strings high in spatial redundancy are recognized more accurately than strings low in such redundancy (Mason 1975, 1978; McClelland, 1976; McClelland & Johnston, 1977; Massaro et al., 1979, 1980). Although there has been some support in the literature for the notion that bigram and trigram frequency influence perceptual processing independent of regularity and spatial redundancy (Massaro, Jastrzembski, & Lucas, 1981; Massaro et al., 1980), such evidence has not been consistently obtained under differing procedures (Gernsbacher, 1984; Gibson, Shurcliff, & Yonas, 1970; Johnston, 1978; Manelis, 1974; McClelland & Johnston, 1977).

The question of central interest to the present paper is whether sensitivity to structural constraints of the orthography is related to speech production. One suggestion is that this sensitivity is acquired through experience with how the orthography maps the spoken language. For example, Gibson et al. (1962) suggested that experience with a consistent mapping of letter clusters to pronunciation may aid the reader in acquiring an appreciation of orthographic structure. Related to this notion, Venezky and Massaro (1979) suggested that phonics instruction, with its emphasis on analytic reading through attention to regular spelling-pronunciation correspondences, may help the beginning reader to acquire information about allowable letter sequences. In contrast to the importance that such suggestions place on a mapping between print and the spoken language, there is the suggestion that a sensitivity to orthographic structure might be acquired through strictly visual means, without reference to the spoken language (e.g., Baron & Thurston, 1973; Gibson et al., 1970; Mason, 1978). Since structural constraints on the orthography, both linguistic regularities and statistical redundancies, impose recurrent visual patterns, such a suggestion is quite feasible.

One argument that has often been used to support the notion of acquisition via visual means is the finding by several researchers that deaf subjects are sensitive to orthographic structure in word recognition and spelling (Dodd, 1980; Doehring & Rosenstein, 1960; Gibson et al., 1970; Hanson, 1982b; Hanson, Shankweiler, & Fischer, 1983; Stone, 1980). It is often assumed that deaf subjects could not employ mapping between written and spoken language, and that the orthographic structure effect must therefore be purely visual (see, for example, Baron & Thurston, 1973; Gibson et al., 1970). As some have noted earlier however, such a conclusion need not necessarily follow (see, for example, Coltheart, 1977; Crowder, 1982). As a rule, deaf children in English-speaking countries receive intensive instruction in speaking and lipreading; this is true both in schools that use an oral educational approach (speech being the means of communication in the classroom) and in schools that use a simultaneous or total communication approach (with speech being accompanied by manual communication in the classroom). Through this speech training, some prelingually, profoundly deaf persons develop quite good speech skills; others develop very little. In between these two extremes, there exists a continuum. Thus, the findings that deaf subjects display a sensitivity to orthographic structure does not necessarily imply a purely visual basis.

The studies examining deaf subjects' sensitivity to orthographic structure have not discriminated between whether the benefit obtained for orthographic structure was due to structure based on orthographic regularity or statistical redundancy. The only attempt to do so was by Gibson et al. (1970). Using multiple regression analyses, they found that sequential redundancies contributed only minimally to performance in a tachistoscopic full report task, and was no greater a predictor of performance for deaf subjects than for hearing subjects. However, since Gibson et al. (1970) did not control for word length, it has been suggested that their study may not be an adequate test of the statistical redundancy descriptions of orthographic structure (Massaro et al., 1980, 1981).

Nor have any of the studies examining deaf subjects' sensitivity to orthographic structure examined how such sensitivity might vary in relation to subjects' speech skills. Although Gibson et al. (1970) found that the number of errors in their letter recall task was not related to speech intelligibility, these investigators did not examine whether the magnitude of any orthographic structure effects varied as a function of speech skills.

The present study examines sensitivity to orthographic structure among two groups of deaf subjects: those with relatively good speech productions, and those with poor speech productions. Their performance will be compared with that of a control group of hearing subjects in two tasks: 1) a perceptual task and 2) a judgment task that examines the extent to which subjects in the three groups are influenced by orthographic structure in rating how word-like certain letter strings appear. To determine the degree to which subjects are sensitive to orthographic regularity and to positional redundancy, these two types of structure are independently varied in the stimuli of the two tasks. If sensitivity to linguistically-based orthographic regularities is related to expertise in speech, then deaf readers with poor speech skills may have difficulty in using orthographic structure, while deaf readers with fairly good speech skills would be expected to exhibit little or no difficulty in using this type of structure. However, the fact that orthographic regularity, by definition, is based on phonological constraints

does not necessarily mean that the reader need be aware of these constraints in order to appreciate such regularity. If the principles of regularity can be acquired from visual patterns, then deaf readers, regardless of their speech skills, would be expected to be as sensitive as hearing readers to these regularities. Since statistical redundancy measures are based on visual properties inherent in the written representation of words, such structure is a feature of the orthography that might be expected to be as readily accessible by deaf readers, regardless of their speech skills, as by hearing readers. Spatial (positional) redundancy is the measure of statistical redundancy tested here. By this measure, the frequency of a letter string is based on the sum of the frequency for each letter in the string at its position of occurrence (Mason, 1975). The frequency of each letter in this summed single letter positional frequency measure is taken from the Mayzner and Tresselt (1965) letter frequency counts.

Method

Subjects

Subjects for the study were two groups of deaf subjects and a control group of hearing subjects. The two groups of deaf subjects differed in the intelligibility of their speech productions: One group had relatively good speech, the other had relatively poor speech. All were paid volunteers.

Deaf subjects. The deaf subjects were prelingually, profoundly deaf. They were undergraduates or recent graduates of Gallaudet College, a liberal arts college for deaf students. All were experienced signers. Background information on hearing loss and speech intelligibility ratings for each of the subjects was obtained from school records.

The two deaf subject groups were determined on the basis of the speech intelligibility ratings of the subjects. These ratings were judgments made by experienced listeners on the staff of the college. In making these judgments, the listeners heard a tape recording of each student's reading of a passage, and were asked to rate, on a scale of 1 - 5, the intelligibility of the student's speech. A '1' on the scale represents speech that is readily understood by the general public, a '5' represents speech that cannot be understood by listening to the tape.

For the purposes of this experiment, the good speech group was defined as subjects who had a speech intelligibility rating of 1, 2, or 3 and the poor speech group was defined as those subjects who had a rating of 4 or 5. There were 11 subjects in the good speech group, and 12 in the poor. The data of three of these subjects were eliminated from analysis: In one case (a subject in the good speech group) the subject failed to meet the accuracy criterion for inclusion in the experiment, and in the other two cases (subjects in the poor speech group) the data of the subjects were lost owing to equipment problems. As a result, there were 10 subjects in each of the two deaf groups.

There were no audiological conditions that readily distinguished between deaf subjects in the two groups. The subjects in the good speech group had a median hearing loss of 100.5 dB (Range = 83-113), better ear average. The subjects in the poor speech group had a median hearing loss of 103 dB (Range = 90-113), better ear average. Measures of residual hearing and vowel discrimination were available for six of the subjects in the good speech group

and for eight of the subjects in the poor speech group. Since response/no response in the frequency of 2,000 Hz and above has been found to be related to speech intelligibility (Smith, 1972), the measure of residual hearing used here was whether or not there was a response at 2,000 Hz or above in the better ear. Three of the subjects in the good speech group and six in the poor speech group did have responses in this range. In terms of vowel discrimination (better ear), the median discrimination of the subjects in the good speech group was 40.0% (Range = 24-76%) and in the poor speech group was 32.5% (Range = 0-52%). For five of the ten subjects in each group, the presence of deafness in immediate family members (parents and/or siblings) suggested that the etiology of deafness was hereditary.

Hearing subjects. The hearing subjects were 17 college undergraduates or recent graduates from the New Haven, Connecticut, area (primarily from Yale University). All had normal hearing and were native speakers of English. The data of five of these subjects were eliminated from analysis: one owing to equipment failure, and four owing to accuracy outside the acceptable range. This resulted in twelve subjects in the hearing group.

Stimuli

The experimental stimuli were the six-letter nonsense words from List 1 of Massaro et al. (1979). These stimuli were constructed to vary orthographic regularity and letter positional frequency independently. This resulted in four types of stimuli: strings high in summed single letter positional frequency that were orthographically regular (e.g., REMOND, SIFLET) or irregular (e.g., RMNOED, TLFIES) as well as strings low in summed positional frequency that were regular (e.g., ENDROM, ESTFIL) or irregular (e.g., RDENMO, EFLSTI). Forty words of each type were included in the experimental list. The same stimuli were used in both the perceptual task and the judgment task.

Procedure

A perceptual task and a judgment task, similar to those in earlier studies testing hearing subjects (e.g, Massaro et al., 1979, 1980), were administered to each of the subjects. The inclusion of the hearing subjects in the present study allowed for a replication of the earlier studies under the present test conditions. In addition to these tasks, a Reading Test was given to obtain a measure of each subject's reading achievement level.

Perceptual task. Subjects were told that they would be seeing letter strings that were word-like but were not actual words. After each string, a probe letter would appear. If that probe letter was present in the string they just saw, they were to press a right-hand button to indicate the response YES. If the probe letter was not present, they were to press a left-hand button to indicate the response NO. There were no time constraints on responding. Subjects were informed that each letter string would be shown for just a brief time and that the length of presentation would be adjusted throughout the task to maintain the accuracy rate at about 75%. In addition, they were informed that half the trials would have the probe letter present, while the other half would not, and that they should therefore have about half YES responses and about half NO responses. For the deaf subjects, instructions were signed in American Sign Language (ASL) by a deaf experimenter, a native signer of the language. For the hearing subjects, instructions were spoken by a hearing experimenter.

Stimuli were displayed for a controlled duration in the center of a CRT display driven by an Atari microcomputer. Following stimulus presentation, a non-character dot mask was presented for 250 ms. Following offset of the mask, a probe letter was presented 3 spaces to the left of the stimulus item, on the same line. This probe remained on until the subject responded. There was an intertrial interval of 250 ms. Since the uppercase character set of the Atari was clearer than the lowercase character set, the stimuli were presented in all uppercase letters. The four stimulus types were mixed throughout each block.

As practice, subjects were presented with 20 blocks of 8 trials each. Following each practice block, the percentage accuracy on the block was displayed. The initial exposure duration was set at 325 ms. Based on the accuracy at the end of each block, the exposure duration was adjusted in steps of 10-25 ms to be longer or shorter to attain 75% accuracy. Practice trials were taken from Massaro et al. (1979), List 2.

Each letter string was used once as a target trial (i.e, the probe letter was present in the strings) and once as a catch trial (i.e., the probe letter was not present in the string). These experimental stimuli were presented in 4 blocks of 80 trials each. Each of the subjects was tested with a randomly-chosen ordering of these four test blocks. Following each block, exposure duration was adjusted, if necessary, to maintain approximately 75% accuracy. The criterion for inclusion of subjects in the study was accuracy within the range of 60-90%. The mean exposure durations were 164.7 ms (SD = 44.1) for the ten deaf subjects in the good speech group, 155.7 ms (SD = 35.9) for the ten deaf subjects in the poor speech group, and 125.0 ms (SD = 42.1) for the twelve hearing subjects. This difference in exposure durations for the three subject groups was not statistically significant, $\underline{F}(2,29) = 2.89$, $\underline{p} > .05$.

Judgment task. Following the perceptual task, the judgment task was administered. The stimuli were typed, in a random order, in uppercase letters on pages of 40 stimuli each. Following each string was a line on which subjects were to indicate their rating. The four test pages were presented in a randomly-chosen order for each of the subjects.

Written instructions informed subjects that their task was to rate several letter strings in terms of how "word-like" the strings were. The instructions indicated that none of the strings were real English words, but that some of the letter strings might seem more "word-like" than other strings. Subjects were shown a drawing of a scale from 1-10 with the numbers equally spaced and were told to use this scale for their ratings, with the number 1 marked as the "worst," being not much like an English word, and the number 10 marked as the "best," being very much like an English word. They were instructed to use all the numbers from 1-10, and to look quickly through the whole set of stimuli before starting to write down their ratings.

One deaf subject in the good speech group, owing to time considerations, was not given the judgment task. The data of one hearing subject were excluded from this analysis as the person failed to use the rating scale correctly. (This hearing subject used the numbers 0 through 10 rather than the numbers 1 through 10, as instructed.)

Reading test. The comprehension subtest of the Gates-MacGinitie Reading Test (1969, Survey F, Form 2) was administered to all subjects. Form F is designed to be appropriate to hearing students in grades 10 though 12, a level that, based on the author's past research, was deemed appropriate for the deaf subjects. A score for reading achievement of each subject was a standard score based on the grade equivalent of 10.1. By this standard score, a score of 50 represents reading achievement of grade 10.1 and each ten points represents performance that is one standard deviation better or worse than grade 10.1.

Results

Perceptual Task

The results of the perceptual task will be considered first. A 3 X 2 X 2 X 2 analysis of variance was performed on the percent correct responses in this task for the three groups of subjects with regularity (regular, irregular), summed positional frequency (high, low), and trial type (target, catch) varied within subjects. The same effects were significant whether the data were subjected to an arcsine transformation or were untransformed. The results reported here are for the untransformed data. The analysis revealed a significant main effect of orthographic regularity, $F(1,29) = 54.41$, $p < .001$, that was qualified by an interaction with group, $F(2,29) = 3.93$, $p < .05$. As shown in Figure 1, this interaction resulted from the deaf subjects in the poor speech group demonstrating less of an advantage due to orthographic regularity than the subjects in the other two groups. Hearing subjects were 7.4% more accurate for regular than irregular letter strings and deaf subjects in the good speech group were 7.0% more accurate for regular than irregular strings. In contrast, deaf subjects in the poor speech group were only 2.6% more accurate for regular strings. (Although this regularity advantage for the deaf subjects in the poor speech group was small, it was still significant, $F(1,9) = 5.52$, $p < .05$, as determined in a post hoc analysis.) There was also a significant main effect of frequency, $F(1,29) = 19.60$, $p < .001$, that did not interact with subject group, $F < 1$. Overall, subjects in the three groups were 4.0% more accurate for high than low frequency strings. There were two significant three-way interactions involving regularity X trial type. The first was the interaction of these two factors with frequency, $F(1,29) = 5.18$, $p < .05$, reflecting greater facilitation due to regularity for high than low frequency strings in the target trials, but a greater effect of regularity for low frequency strings in the catch trials. The second was the interaction of these two factors with subject group, $F(2,29) = 3.87$, $p < .05$, reflecting greater facilitation due to regularity on target than catch trials for the hearing subjects, but a greater effect of regularity on catch trials than target trials for deaf subjects in the good speech group. The facilitation due to regularity for deaf subjects in the poor speech group was quite small in both cases. The mean percentages correct for each subject group as a function of regularity, frequency, and trial type are given in Table 1.

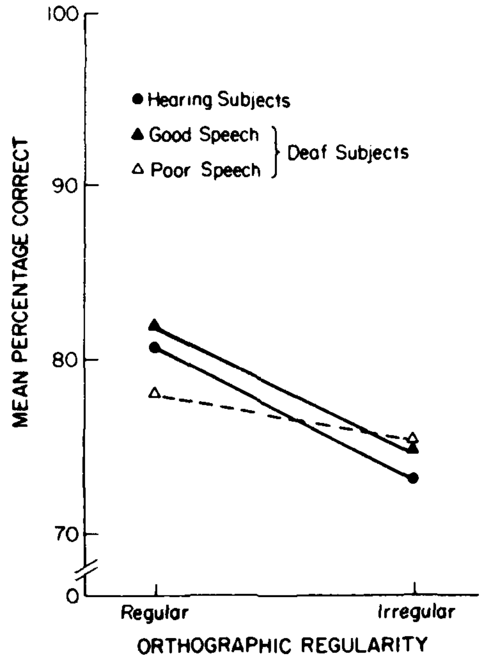


Figure 1. *Mean percentage correct responses as a function of orthographic* regularity for hearing subjects, deaf subjects with good speech, and deaf subjects with poor speech.

Judgment Task

The judgment task was used to determine the extent to which subjects were influenced by orthographic regularity and spatial redundancy in decisions about how word-like letter strings appeared. As shown in Table 2, subjects in all three groups rated orthographically regular strings as more word-like than irregular strings, and rated strings high in single letter positional frequency as more word-like than strings low in such frequency.

An analysis of variance of the ratings data for the factors of subject group X regularity X frequency obtained a main effect of subject group, $F(2,27) = 5.67$, $p < .01$, indicating that there was a difference in absolute ratings between the subject groups. A post hoc analysis indicated that this difference was due to the deaf subjects with good speech generally rating the letters strings as less word-like than subjects in the other two groups (Newman-Keuls, $p < .05$). The mean absolute ratings for subjects in the three groups were 4.79 for the hearing subjects, 4.97 for the deaf subjects with poor speech, and 3.54 for the deaf subjects with good speech. Since the conservative use of the rating scale by the deaf subjects with good speech would have reduced indications of orthographic sensitivity, the ratings of these subjects cannot be fairly compared with those of the subjects in the

Table 1

Mean percentage correct in the perceptual task for each subject group as a function of orthographic regularity (regular, irregular), summed single letter positional frequency (high, low), and trial type (target, catch).

		Target		Catch	
		Regular	Irregular	Regular	Irregular
Hearing	High	83.1	70.8	81.0	80.2
	Low	75.6	66.3	83.1	75.8
Deaf-Good speech	High	84.0	78.9	84.9	75.7
	Low	78.9	74.9	80.9	71.1
Deaf-Poor speech	High	80.3	76.9	78.6	79.9
	Low	77.5	74.0	76.5	71.5

Table 2

Mean ratings in the judgment task for each subject group as a function of orthographic regularity and summed single letter positional frequency.

Frequency	Regularity		
	Regular	Irregular	Mean
Hearing			
High	7.7	3.2	5.4
Low	5.9	2.4	4.1
Mean	6.8	2.8	
Deaf-Good Speech			
High	5.5	2.6	4.1
Low	3.9	2.1	3.0
Mean	4.7	2.4	
Deaf-Poor Speech			
High	6.9	4.7	5.8
Low	5.4	3.0	4.2
Mean	6.1	3.8	

other two groups. Therefore, two different analyses were performed on the ratings data: one on the ratings of the hearing subjects and the deaf subjects in the poor speech group, the second on the ratings of the deaf subjects in the good speech group.

In the first analysis with the two subject groups, there were large main effects of both regularity, $F(1,19) = 257.37$, $p < .001$, and frequency, $F(1,19) = 158.39$, $p < .001$. There was also an interaction of regularity X subject group, $F(1,19) = 18.58$, $p < .001$, reflecting greater effects of regularity for the hearing subjects than the deaf subjects. (A post hoc analysis, however, indicated that the effect of regularity was still significant when only the deaf subjects with poor speech were considered, $F(1,9) = 43.01$, $p < .001$.) The only other effect to approach significance was an interaction of regularity X frequency X group, $F(1,19) = 3.95$, $p < .07$. Post hoc analyses determined that this interaction was due to the fact that for the hearing subjects, but not for the deaf subjects with poor speech, regularity was a much greater determiner of wordness than was frequency (there was a significant interaction of regularity X frequency for the hearing subjects, $F(1,10) = 23.15$, $p < .001$, that was not obtained for the deaf subjects with poor speech, $F < 1$).

In the second analysis, of only the deaf subjects with good speech, there were significant main effects of both regularity, $F(1,8) = 89.03$, $p < .001$, and frequency, $F(1,8) = 93.38$, $p < .001$, as well as an interaction between these variables, $F(1,8) = 44.15$, $p < .001$. This interaction reflected the fact that regularity was a greater determiner of ratings than was frequency.

Correlations of Perceptual and Judgment Data

To examine whether the same factors that influenced perceptual processing also influenced subjects' decisions about how word-like the letter strings were, subjects' ratings in the judgment task were correlated with their accuracy in the perceptual task. A mean percentage correct score was determined for each of the three subject groups in the perceptual task for each of the 160 stimulus items. For the judgment task, a mean rating for each of the 160 stimuli was calculated for each group. Results of the correlations between the two tasks are given in Table 3. Except for the subjects in the poor speech group, analysis of subjects' performance in the two tasks revealed significant correlations between tasks and groups. That is, for the hearing subjects and for the deaf subjects in the good speech group, the more accurately a letter string was responded to in the perceptual task, the more highly word-like it was rated in the judgment task. Moreover, the letter strings that were perceived accurately and rated high were the same for those two subject groups. In contrast, the accuracy performance of the deaf subjects in the poor speech group not only failed to correlate significantly with the ratings of the other two subject groups, but also failed to correlate significantly with their own ratings. Thus, it appears that the sensitivity to orthographic structure measured in the perceptual task was related to such sensitivity measured by the judgment task for the hearing subjects and the deaf subjects with good speech, but not for the deaf subjects with poor speech.

As a means of providing converging information about subjects' sensitivity to orthographic structure, post hoc correlations were undertaken on measures of orthographic structure and subjects' performance on the two

Table 3

Correlations between deaf and hearing subjects' performance in the perceptual and judgment tasks.

	Judgment task		
Perceptual task	Deaf-Good Hearing	Deaf-Poor Speech	Speech
Hearing	.32 *	.29 *	.26 *
Deaf good speech	.30 *	.28 *	.25 *
Deaf Poor speech	.17	.12	.17

Note: * p <.01, df = 158, one-tailed

Table 4

Correlations of subjects' performance in the perceptual and judgment tasks with orthographic regularity and summed single letter positional frequency.

	Regularity	Frequency
Perceptual task		
Hearing	.27 *	.25 *
Deaf-Good speech	.24 *	.26 *
Deaf-Poor speech	.08	.21 *
Judgment task		
Hearing	.83 *	.39 *
Deaf-Good speech	.72 *	.46 *
Deaf-Poor speech	.65 *	.54 *

Note: * p < .01, df = 158, one-tailed.

tasks. The measure of orthographic regularity was the dummy regularity measure of Massaro et al. (1981).[1] According to this measure, each of the 160 stimulus items is assigned the binary classification of '0' if it is orthographically regular, or '1' if it is irregular. The measure of single letter frequency was determined on the basis of the position-sensitive log-frequency tables given in Massaro et al. (1980).[2] For the present stimuli, these two measures were not significantly correlated, $r = .16$, df = 158, $p > .01$, one-tailed).

As can be seen in Table 4, regularity significantly correlated with the performance of the deaf and hearing subjects in the two tasks, with only one exception. The exception, again, was the deaf subjects in the poor speech group on the perceptual task. Consistent with the results of the orthogonal contrasts, the accuracy of the hearing subjects and the deaf subjects in the good speech group in the perceptual task was significantly correlated with orthographic regularity. That is, those subjects were more accurate on regular than irregular strings. The accuracy of the deaf subjects in the poor speech group was not significantly correlated with regularity. In the judgment task, however, the ratings of subjects in all three groups were significantly correlated with regularity, with higher ratings for regular than irregular strings. Single letter positional frequency significantly correlated with the performance of subjects in each of the three groups in the two tasks, as shown in Table 4. In all cases, strings high in frequency were responded to more accurately and rated as more word-like than strings low in frequency. As can be seen, the correlations between performance and frequency in the judgment task were not as high, however, as the correlations between performance and regularity.

As can also be seen in Table 4, the correlations with regularity and frequency were comparable for the deaf and hearing subjects in the perceptual task, with the exception, of course, of the deaf subjects in the poor speech group. However, when the deaf and hearing subjects were compared on the judgment task, a difference between the groups emerged: The correlations with regularity for the deaf subjects with poor speech were significantly less than for the hearing subjects, $t(157) = 7.19$, $p < .001$, two-tailed, whereas the correlations with frequency were significantly <u>greater</u> for the deaf subjects with poor speech than for the hearing subjects, $t(157) = -3.94$, $p < .001$, two-tailed. (Since the deaf subjects in the good speech group demonstrated a conservative use of the rating scale, a restricted range problem was indicated for these subjects. This problem would have tended to reduce the magnitude of the correlations of their ratings data with both regularity and frequency, making comparisons of their correlations with those of subjects in the other two groups difficult to interpret.)

Correlations with Reading Proficiency

Finally, analyses were performed to determine whether sensitivity to structural constraints of the orthography varied as a function of reading proficiency in either task for the deaf subjects. There was nothing in the data to suggest any such relationship. The mean reading score of the deaf subjects in the good speech group was 49.0 and of those in the poor speech group was 46.2. Thus, subjects, on the average, were reading at very nearly 10th grade level, a level indicating that they were quite successful readers by comparison with most prelingually, profoundly deaf individuals (for discussion of reading ability of deaf individuals see, for example, Conrad,

1979, and Karchmer, Milone, & Wolk, 1979). The reading scores of the two groups did not differ significantly, $t < 1$. There were no significant correlations between reading comprehension and the regularity advantage or the frequency advantage on either task (all $ps > .05$, two-tailed).

The hearing subjects were also given the reading test, but their performance could not accurately be ascertained on the scale. The accuracy of many of these subjects was so great that it fell outside the range for which the test had reliable norms. All that can reasonably be reported about the hearing subjects' data is that all of them obtained scores of 70 or greater.

Discussion

Consistent with earlier studies, deaf subjects in the present study were found to be sensitive to orthographic structure (Doehring & Rosenstein, 1960; Gibson et al., 1970; Hanson 1982b; Stone, 1980). Such findings have often been taken as evidence that orthographic sensitivity need not be related to an appreciation of the phonological constraints that govern word formation. That is, since deaf individuals are presumed not to use speech, it follows that if they have acquired a sensitivity to orthographic structure principles then they must have acquired it through strictly visual means, quite independently of experience with how the written language maps the spoken. As mentioned earlier, however, such an interpretation of the findings with deaf subjects is problematic. Deaf individuals generally do have some experience with speech, although they differ in their expertise in this area: some are quite proficient with speech and others are considerably less so. The present study investigated whether sensitivity to two aspects of orthographic structure (namely, orthographic regularity and statistical redundancies) relate to speech intelligibility by comparing the orthographic sensitivity of hearing subjects with that of two groups of deaf subjects who varied in one aspect of speech proficiency--speech intelligibility--but did not differ in their reading proficiency or, in any discernible respect, audiometrically.

The outcome of the perceptual and judgment tasks indicated that sensitivity to orthographic regularity (defined in terms of phonological and scribal constraints) differed as a function of expertise in speech. In the perceptual task, it was found that those deaf subjects with good speech exhibited perceptual facilitation due to regularity that was comparable to that of the hearing subjects. Those deaf subjects in the poor speech group exhibited much less facilitation than those in the other two groups. Post hoc correlations provided additional evidence for this relationship; the accuracy of the deaf subjects in the good speech group, like that of the hearing subjects, was significantly correlated with orthographic regularity, but the accuracy of the deaf subjects in the poor speech group was not. The results of the judgment task were consistent with the perceptual task in indicating a relationship between speech intelligibility and sensitivity to orthographic regularity. In that task, the correlation with regularity was not as great for the subjects with poor speech as for the hearing subjects nor, apparently, for the deaf subjects with good speech.

It is worth noting that the deaf subjects in the poor speech group did not appear to be completely insensitive to orthographic regularity. In the perceptual task, these subjects exhibited a small facilitation due to regularity that was significant in the orthogonal contrast, although it failed to reach significance in the post hoc correlation. Given the significance in

the orthogonal contrast, though, it might be posited that this type of structure does influence their perceptual processing to some limited extent. Moreover, in the judgment task their ratings were significantly higher for regular than irregular strings, and there was a significant correlation between their ratings and the post hoc measure of regularity. This sensitivity to regularity on the part of the deaf subjects with poor speech is not inconsistent with the notion that such sensitivity is related to speech intelligibility. It must be borne in mind that even these readers are not completely without speech ability--their proficiency with speech is just less than that of the hearing subjects and the deaf subjects in the good speech group. Correspondingly, their sensitivity to orthographic regularity was found to be less.

It is of interest that the present study found that the perceptual facilitation of the deaf subjects in the good speech group was comparable to that of the hearing subjects. Although subjects in this group had good speech in relation to other deaf speakers, the speech of most of these subjects was only moderately intelligible. Only three of the subjects in this group had speech that was rated as better than a '3' on the speech intelligibility rating scale (a '3' represents speech that the general public has some difficulty in understanding, at least initially). Thus, sensitivity to orthographic regularity can apparently be acquired without perfect production of speech. What is crucial is not that speech is perfectly intelligible as perceived by listeners, but that the deaf individual is able to appreciate the phonological distinctions of the language. Although some correlation undoubtedly exists between perceived intelligibility and phonological appreciation, the two are not one and the same. The group of deaf subjects in this study whose speech was only moderately intelligible to listeners were, apparently, quite phonologically competent.

In contrast to the indications for regularity, the deaf subjects in both the good and poor speech groups exhibited a sensitivity to spatial (positional) redundancy that was no less than that of the hearing subjects. This finding suggests that these statistical redundancies, which are based on properties of the visual signal itself, can be learned through strictly visual means. The subjects in all three groups were influenced by spatial redundancy information in their ratings, but the deaf subjects with poor speech showed higher correlations with frequency than the hearing subjects. This is suggestive that deaf readers with poor speech may compensate for their lesser proficiency with regularity by relying more heavily on statistical redundancies of the orthography.

The difference in sensitivity to orthographic regularity as a function of speech intelligibility stands as the major finding of the present study, suggesting an important relationship between expertise in speech and acquisition of orthographic regularity (e.g., Gibson et al., 1962; Venezky & Massaro, 1979). Given the correlational nature of this finding, however, it cannot be determined from this study how regularity and speech intelligibility are causally linked. One possibility is that direct relationships between sensitivity to orthographic regularity and speech exist. For example, it could be that speech ability improves an individual's ability to perform a linguistic analysis of words, an analysis that would provide the information needed to acquire an appreciation of the phonological structure of words underlying orthographic regularity.

Alternatively, it is possible that the tasks of the present study tapped the use of an internal speech code, and that the obtained relationship between orthographic regularity and speech intelligibility reflects the fact that both are related to this internal code. In this regard, the present findings are compatible with results from short-term memory studies. In those studies, hearing readers have been more effectively able than deaf readers to use a speech code, and deaf readers with good speech intelligibility have been more effectively able than deaf readers with poor speech intelligibility to use a speech code (Conrad, 1979; Hanson, 1982a; Lichtenstein, in press). The obtained relationship is generally assumed to be causative, such that the better speech skills promote ability to use an internal speech code (see Conrad, 1979).

In actuality, other factors (e.g., lipreading and reading achievement) also have been found to be associated with the ability to use an internal speech code by deaf readers (Conrad, 1979; Lichtenstein, in press). It is likely that there is no simple relationship among these factors; probably there are multiple directions of causation. For example, good speech production could promote acquisition of an internal speech code, which, in turn, could promote lipreading skill. This lipreading skill could then serve to sharpen the speech code, which could then further enhance speech production. Similarly with reading, an effective speech code could promote reading success, and experience with reading could provide information that would serve to enhance the internal code, lipreading, and speech production. Such interactions between language forms need not be limited to deaf readers. These same factors could also interact for hearing individuals in the acquisition of linguistic sensitivity, although hearing readers would have the advantage of an additional reliable auditory input.

In addition to the factors named above, another source of linguistic input might influence acquisition of linguistic sensitivity for deaf readers: for deaf readers skilled in manual communication, fingerspelling could prove useful. Fingerspelling is a manual communication system in which words are spelled out by the sequential production of the handshapes of a manual alphabet. (The American manual alphabet uses a one-handed configuration for each letter; the British system uses a two-handed configuration for each letter.) For deaf persons skilled in fingerspelling, orthographically permissible letter strings conform to the structure inherent in the manual production. As a result, production of illegal letter strings would feel "difficult" or "awkward" to produce on the hand. Thus, it is reasonable to hypothesize that fingerspelling could be useful in acquisition of orthographic structure. While fingerspelling may contribute, in part, to sensitivity to orthographic structure for deaf readers, since the deaf subjects in both groups were skilled signers, the observed differences in sensitivity between the two groups cannot be accounted for on the basis of fingerspelling.

Although it has been suggested in the literature that hearing children (sixth graders) who are good readers may be more sensitive to both orthographic regularity and spatial redundancy information than are children who are poor readers (Mason & Katz, 1976; Massaro & Taylor, 1980), the same characterization does not appear to distinguish between good and poor hearing readers at the college level (Massaro & Taylor, 1980). In the present study, the deaf readers were less proficient readers than the hearing subjects. Yet, consistent with the earlier findings with hearing college students, no difference in perceptual facilitation due to orthographic structure resulted

from this discrepant reading proficiency. In their perceptual facilitation due to orthographic regularities, the deaf subjects with good speech were comparable to the hearing subjects, and in their perceptual facilitation due to spatial redundancy, the deaf subjects, regardless of their speech production ability, were no less sensitive than the hearing subjects. Moreover, considering only the deaf subjects, advantages due to regularity and spatial redundancy did not correlate significantly with reading comprehension in the perceptual or judgment tasks.

In summary, the present results suggest a relationship between a sensitivity to at least one aspect of orthographic structure, namely, linguistically-based regularity, and expertise in speech. However, sensitivity to spatial redundancy does not appear to be related to such expertise. Further, the present results indicate that despite the fact that regularity and spatial frequency are normally confounded in written English (e.g., Massaro et al., 1980), acquisition of sensitivity to the two can occur independently: Although the deaf subjects with poor speech were less sensitive to regularity than hearing subjects, they were no less sensitive to spatial frequency.

References

Adams, M. J. (1979). Models of word recognition. Cognitive Psychology, 11, 133-176.

Aderman, D., & Smith, E. E. (1971). Expectancy as a determinant of functional units in perceptual recognition. Cognitive Psychology, 2, 117-129.

Baron, J., & Thurston, I. (1973). An analysis of the word-superiority effect. Cognitive Psychology, 4, 207-228.

Carr, T. H., Posner, M. I., Pollatsek, A., & Snyder, C. R. R. (1979). Orthography and familiarity effects in word processing. Journal of Experimental Psychology: General, 108, 389-414.

Coltheart, M. (1977). Critical notice. Quarterly Journal of Experimental Psychology, 29, 157-167.

Conrad, R. (1979). The deaf school child. London: Harper & Row.

Crowder, R. G. (1982). The psychology of reading: An introduction. New York: Oxford University Press.

Dodd, B. (1980). The spelling abilities of profoundly pre-lingually deaf children. In U. Frith (Ed.), Cognitive processes in spelling (pp. 423-440). London: Academic Press.

Doehring, D. G., & Rosenstein, J. (1960). Visual word recognition by deaf and hearing children. Journal of Speech and Hearing Research, 3, 320-326.

Gates-MacGinitie reading tests. (1969). New York: Columbia Teachers' College Press.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. Journal of Experimental Psychology: General, 113, 256-281.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. (1962). The role of grapheme-phoneme correspondence in the perception of words. American Journal of Psychology, 75, 554-570.

Gibson, E. J., Shurcliff, A., & Yonas, A. (1970). Utilization of spelling patterns by deaf and hearing subjects. In H. Levin & J. P. Williams (Eds.), Basic studies in reading (pp. 51-73). New York: Basic Books, Inc.

Hanson, V. L. (1982a). Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall. Journal of Experimental Psychology: Learning, Memory and Cognition, 8, 572-583.

Hanson, V. L. (1982b). Use of orthographic structure by deaf adults: Recognition of fingerspelled words. Applied Psycholinguistics, 3, 343-356.

Hanson, V. L., Shankweiler, D., & Fischer, F. W. (1983). Determinants of spelling ability in deaf and hearing adults: Access to linguistic structure. Cognition, 14, 323-344.

Henderson, L. (1982). Orthography and word recognition in reading. London: Academic Press.

Johnston, J. C. (1978). A test of the sophisticated guessing theory of word perception. Cognitive Psychology, 10, 123-153.

Karchmer, M. A., Milone, M. N., Jr., & Wolk, S. (1979). Educational significance of hearing loss at three levels of severity. American Annals of the Deaf, 124, 97-109.

Lichtenstein, E. (in press). The relationships between reading processes and English skills of deaf college students: Parts I and II. Applied Psycholinguistics.

Manelis, L. (1974). The effect of meaningfulness on tachistoscopic word perception. Perception & Psychophysics, 16, 182-192.

Mason, M. (1975). Reading ability and letter search time: Effects of orthographic structure defined by single-letter positional frequency. Journal of Experimental Psychology: General, 104, 146-166.

Mason, M. (1978). The role of spatial redundancy in grapheme recognition: Perception or inference? Journal of Experimental Psychology: Human Perception and Performance, 4, 662-673.

Mason, M., & Katz, L. (1976). Visual processing of nonlinguistic strings: Redundancy effects and reading ability. Journal of Experimental Psychology: General, 105, 338-348.

Massaro, D. W., Jastrzembski, J. E., & Lucas, P. A. (1981). Frequency, orthographic regularity, and lexical status in letter and word perception. In G. Bower & A. R. Lang (Eds.), The psychology of learning and motivation (Vol. 15, pp. 163-200). New York, Academic Press.

Massaro, D. W., & Taylor, G. A. (1980). Reading ability and utilization of orthographic structure in reading. Journal of Educational Psychology, 1980, 72, 730-742.

Massaro, D. W., Taylor, G. A., Venezky, R. L., Jastrzembski, J. E., & Lucas, P. A. (1980). Letter and word perception: Orthographic structure and visual processing in reading. Amsterdam: North-Holland.

Massaro, D. W., Venezky, R. L., & Taylor, G. A. (1979). Orthographic regularity, positional frequency, and visual processing of letter strings. Journal of Experimental Psychology: General, 108, 107-124.

Mayzner, M. S., & Tresselt, M. E. (1965). Tables of single-letter and digram frequency counts for various word-length and letter position combinations. Psychonomic Monograph Supplements, 1(2), 13-32.

McClelland, J. L. (1976). Preliminary letter identification in the perception of words and nonwords. Journal of Experimental Psychology: Human Perception and Performance, 2, 80-91.

McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. Perception & Psychophysics, 22, 249-261.

Singer, M. H. (1980). The primacy of visual information in the analysis of letter strings. Perception & Psychophysics, 27, 153-162.

Smith, C. R. (1972). Residual hearing and speech production in deaf

children. Doctoral dissertation, The City University of New York.

Spoehr, K. T., & Smith, E. E. (1975). The role of orthographic and phonotactic rules in perceiving letter patterns. Journal of Experimental Psychology: Human Perception and Performance, 1, 21-34.

Stone, R. E. (1980). Factors affecting the tachistoscopic recognition of trigrams for both deaf and hearing college students. Proceedings of the Gallaudet Conference on Reading in Relation to Deafness (pp. 171-192). Washington, DC: Division of Research, Gallaudet College.

Venezky, R. L., & Massaro, D. W. (1979). The role of orthographic regularity in word recognition. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1, pp.85-107). Hillsdale, NJ: Erlbaum.

Footnotes

[1]Alternatively, a measure of orthographic regularity in terms of an irregularity count is possible (see Massaro et al., 1980, 1981). The present data were also analyzed using the irregularity count measure described in Table II of Massaro et al. (1981). The dummy measure, however, proved to discriminate better between the three subject groups than did the irregularity count. Therefore, the results reported here are for the dummy regularity measure.

[2]Post hoc correlations with bigram and trigram frequency are also possible, and such measures have been found to correlate highly with accuracy on tests of perceptual facilitation in other studies (Massaro et al., 1980, 1981). However, these measures correlate very highly with orthographic regularity (Massaro et al., 1980). Therefore, post hoc correlations with these frequency measures are not considered here.

COOPERATIVE PHENOMENA IN BIOLOGICAL MOTION*

J. A. S. Kelso† and J. P. Scholz†

1. Introduction

The production of a "simple" utterance, such as the syllable /ba/, involves the cooperation of a large number of neuromuscular elements operating on different time scales, e.g., at respiratory, laryngeal, and supralaryngeal levels. Yet somehow, from this huge dimensionality, /ba/ emerges as a coherent and well-formed pattern. Similarly, were one to count the neurons, muscles, and joints that cooperate to produce the "simple" act of walking, literally thousands of degrees of freedom would be involved. Yet again, somehow walking emerges as a fundamentally low-dimensional cyclical pattern--in the language of dynamical systems, a periodic attractor. In physics, an infinite dimensional system, described by a complicated set of partial, nonlinear differential equations can be reduced--when probed experimentally or analyzed theoretically--to a low-dimensional description (Procaccia, this volume;* Shaw, 1981). In all these cases, it seems, information about the system is compressed--from a microscopic basis of huge dimensionality--to a macroscopic basis of low dimensionality.

Our particular interest is how such compression occurs in the multidegree of freedom actions of people and animals. How does an internally complex system "simulate" a simpler, lower dimensional system? As we shall see, an important feature of our efforts to understand the control and coordination of movement is the concept of order parameter (Haken, 1975, 1983; see also Kelso & Tuller, 1984). Order parameters define the collective behavior of the system's many components in terms of its essential variables alone; they are few in number even in very complicated physical and chemical systems. Note how the emphasis on discovering order parameters takes us away from a focus on individual elements (regardless of the level at which these elements are described): Just as the motion of a single molecule is not relevant to the essential description of the behavior of a gas, so too, one suspects, the action of a single reflex is not relevant to the essential description of an organism's behavior.

---

*In H. Haken (Ed.), Synergetics of complex systems: Operational principles in neurobiology, physical systems, and computers. Springer-Verlag, 1985.

†Also University of Connecticut.

Acknowledgment. Research supported by ONR Contract N00014-83-C-0083, NINCDS Grant NS-13617, and BRSG Grant RR-05596. We are very grateful to Professor H. Haken for his continued interest and encouraging comments on this work, as well as his coworkers in Stuttgart (particularly G. Schöner and H. Bunz) for their helpful, informed discussions.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]

Our focus here is on the spatiotemporal patterns formed by the ensemble activity of neurons, muscles, and joints during the performance of a coordinated act. As Weisskopf (1984) emphasizes in a different context, such problems rest with defining relations between different aggregates of atoms or molecules, and of the modes of transition from one structure to another. The abstraction of a system's order parameters is thus of paramount importance, because it allows one to separate the essential from the nonessential, thereby enabling a complex phenomenon to become more transparent. This "macroscopic" strategy is brought to bear here on our efforts to discover the principles underlying the control and coordination of movements. In the following sections, we first briefly summarize evidence for the existence of unitary processes in complex actions and describe some of the characteristic properties of such units. From such analysis, the phase relation among the motions of skeletomuscular components will emerge as a candidate order parameter. We then contrast various theoretical notions about pattern generation in movement and introduce some recent evidence in favor of a synergetic approach. Synergetics motivates the treatment of complicated biological motion as fundamentally a cooperative phenomenon. In support of this view, certain kinds of activities will be shown to display the features of a nonequilibrium phase transition.

2. A Unitary Process (Coordinative Structure)

For the Soviet physiologist Bernstein (1967), the existence of a large number of potential degrees of freedom in the motor system precluded the possibility that each was controlled individually at every point in time. Rather, he hypothesized that the central nervous system (CNS) "collects" multiple degrees of freedom into functional units that then behave, from the perspective of control, as a single degree of freedom. During a movement, the internal degrees of freedom are not controlled directly, but are constrained to relate among themselves in a relatively fixed and autonomous fashion. But is it, in fact, the case that in coordinated actions, the many neuromuscular components actually function as a single degree of freedom?

Support for the hypothesis that a group of relatively independent muscles and joints forms a single functional unit would be obtained if it were shown that a challenge or perturbation to one or more members of the group was, during the course of activity, responded to by other remote (nonmechanically linked) members of the group. We have recently found that speech articulators (lips, tongue, jaw) produce functionally specific, near-immediate compensation to unexpected perturbation, on the first occurrence, at sites remote from the locus of perturbation (Kelso, Tuller, V.-Bateson, & Fowler, 1984). The responses observed were specific to the actual speech act being performed: for example, when the jaw was suddenly perturbed while saying the syllable /bæb/, the lips compensated so as to produce the final /b/, but no compensation was seen in the tongue. Conversely, the same perturbation applied during the utterance /bæz/ evoked rapid and increased tongue muscle activity (so that the appropriate tongue-palate configuration for a fricative sound was achieved), but no active lip compensation.

Recent work has also varied the phase of the jaw perturbation during bilabial consonant production. Remote reactions in the upper lip were observed only when the jaw was perturbed during the closing phase of the motion, that is, when the reactions were necessary to preserve the identity of the spoken utterance. Thus the form of cooperation observed is not rigid or

"hard wired": the unitary process is flexibly assembled to perform specific functions (for additional evidence in other activities, see Kelso et al., 1984). Elsewhere we have drawn parallels between these findings and brain function in general (Kelso & Tuller, 1984). Just as groups of cells, not single cells, are the main units of selection in higher brain function (Edelman & Mountcastle, 1978), so too task-specific ensembles of neuromuscular elements appear to be the significant units of control and coordination of action.

Stunning evidence attesting to this self-organizational style of neural and behavioral function comes from recent microelectrode studies of somatosensory cortex in adult squirrel and owl monkeys by Merzenich and colleagues (see Merzenich & Kaas, 1984, for review): when the middle finger of the monkey's hand was surgically removed, brain regions representing the other adjacent fingers progressively shifted (over the course of a few weeks) into the missing finger's hitherto exclusive brain region. Also, if a portion of cerebral cortex was injured, the appropriate somatosensory "map" moved to the region surrounding it--a spatial shift of nerve cell activity as it were. These data challenge a view of neural functioning that is determined by "hard-wired" or "fixed" anatomic connections established before or shortly after birth. Just as we have observed rapid "soft" forms of compensation in speech production, so it seems, the brain has a functionally fluid, self-organizing character that allows longer-term compensation for injury.

3. Characteristic Properties of a Unitary Process

A main way to uncover the intrinsic properties of a functional unit of action is to transform the unit as a whole (e.g., by scaling on movement rate, amplitude, etc.) and search for what remains *invariant across transformation*. The discovery of such "relational invariants" (e.g., Kelso, 1981). could provide a useful step toward explicating the design logic of the motor system.

Much evidence now exists from a wide variety of movement activities that <u>relative timing</u> among muscles and kinematic components is preserved across scalar changes in force or rate of production. For example, when a cat's speed of locomotion increases, the duration of the "step cycle" decreases (Grillner, 1975; Shik & Orlovskii, 1976) and an increase in activity is evident in the extensor muscles during the end of the support phase of the individual limb. Notably, this increase in muscle activity (and corresponding development of propulsive force) does not alter the relative timing among functionally linked extensor muscles, although the duration of their activity may change markedly (see Grillner, 1975; Shik & Orlovskii, 1976, for reviews).

Interestingly, there is some limited evidence that this style of organization applies also to speech production. What makes a word a word in spite of differences among speakers, dialects, intonation patterns, and so on? Our view is that the key to this question lies in understanding how the coordinated movements of the vocal tract articulators structure sound for a listener. According to this view, the invariance that allows us to perceive the sounds of a language in so many different contexts exists in the functionally-defined behavior of the articulatory system. But how is such behavior to be described? It is well known, for instance, that the <u>same</u> word has markedly different kinematic, electromyographic, and acoustic attributes when produced in different contexts. A solution to this dilemma may lie in the finding by Tuller, Kelso, and Harris (1982) that the <u>relative timing</u> of

activity in various articulatory muscles is preserved across the very substantial metrical changes in duration and amplitude of muscle activity that occur when a speaker varies his/her speaking rate and stress pattern (for evidence in other motor skills see Shapiro & Schmidt, 1982). An important extension of these earlier EMG findings is the discovery that the relative timing of articulator movements is stable across different speaking rate and stress patterns. Presently, these results apply to the cooperative relations among lips, tongue, jaw, and larynx (see Tuller & Kelso, 1984, for review).

How is the relative timing invariant to be rationalized? A popular view is that time is metered out by a central motor program (see below) that instructs the articulators when to move, how far to move, and for how long. A reconceptualization and consequent reanalysis of the Tuller and Kelso (1984) data, however, strongly suggests that time, per se, is not directly controlled. Using phase plane techniques to represent the motions geometrically, we have shown that critical phase angles--relating one articulator's position-velocity ($x,\dot{x}$) state to another--appear to be most crucial for orchestrating the coordination among articulators (Kelso & Tuller, 1985, in press). The beauty of this gestural phase analysis (which is autonomous and does not require an explicit representation of time) is that it provides a topological description of articulatory behavior that remains unaltered across manifold speaker characteristics. Moreover, critical phase angles are revealed by the flow of the dynamics of the system, not externally defined. Thus, they can serve as natural sources of information for guaranteeing the stability of coordination in the face of scalar (metrical) change (for more details, see Kelso & Tuller, in press).

Finally, there is a strong hint that phase constancy reflects an evolutionary design principle. From the invertebrates, in which many groups employ large numbers of propulsive structures (limbs, tube feet, or cilia) for swimming and locomotion, to the vertebrates that walk, run, or jump using one, two, three, or four pairs of legs, the same design property is apparent, namely, all of these creatures possess processes that communicate information about the phase of activity among component structures (von Holst, 1937/1973; Sleigh & Barlow, 1980). We will develop in more detail below the notion that phase is an essential parameter of complex, coordinated action. We emphasize at this point that a phase constancy indicates a functional constraint on movement, what we call a coordinative structure or unit of action (cf. Easton, 1972; Fowler, 1977; Kelso, Southard, & Goodman, 1979; Turvey, 1977). Thus, during an activity the spatiotemporal behavior of individual components is constrained within a particular relationship. Flexibility can then be attained by adjusting control parameters over the entire unit.

4. Theories of Pattern Generation

The core idea expressed in Sections 2 and 3 above--that a system possessing a large number of potential degrees of freedom is compressed into a single functional unit of action (or coordinative structure) that requires few control decisions--is unorthodox. It differs in significant ways from more conventional treatments of movement based either on the information processing notion of a motor program or the neurally-based notion of a central pattern generator. The motor program, by definition, is an internal representation of a movement pattern that is prestructured in advance of the movement itself. Analogous with a computer program, it constitutes a prescribed set of instructions to the skeletomuscular system. In MacKay's (1980) analysis of a

dynamic activity, the locomotory step cycle, the many kinematic details are ordered a priori by a sequence of commands/instructions to the skeletomuscular apparatus whose role is to implement these instructions. The format of the program is that of a formal machine; symbol strings are employed to achieve (or explain) the order and regularity of the step cycle. As in most programming accounts, the control prescription is highly detailed and the role that dynamics plays in fashioning the pattern is ignored. So also is the interface between the small-scale "informational" contents of the program and the large-scale, energetic requirements of the muscle-joint system. Finally, the contents of the program are not rationalized: a principled basis for selecting desired quantities (e.g., apply flexion torque for 100 ms) is omitted.

The neural counterpart of the motor program is the central pattern generator (CPG). Here too, the order and regularity observed in the world is attributed to a device inside the CNS (a neural circuit) that, when activated, coordinates the different muscles to produce movement (Grillner, 1985). Though subject to feedback influences, the circuit is "hard-wired" and the goal of neuroscience is to locate the neurons that constitute the network and to define their properties and interrelations. Though an admirable enterprise, there are questions about its propriety. For example, the parameter space of a CPG, e.g., the membrane properties of its elements, synaptic connections, etc., has been variously estimated to be 46 or 55 (compare Bullock, 1976, to Bullock, 1980; also Selverston, 1980). Presumably not all of these parameters are necessary to understand a CPG, but principles beyond those of neurophysiology are surely needed to guide the selection of relevant parameters in such a high-dimensional space. As Loeb and Marks (1980) emphasize, principles of operation constitute the knowledge for understanding a CPG and these are disembodied from the actual device (or its model). In addition, even if all the details of a putative CPG were known, the problem of relating the known microproperties to characteristic macroproperties such as the amplitude, phase, and frequency of a wing beat or a step cycle would still remain.

The question then is this: where do the necessary principles come from? For some years now, we have advocated an approach in which problems of biological motion are treated in a manner continuous with cooperative phenomena in other physical, chemical, and biological systems, i.e., as synergetic or dissipative structures (Kelso & Tuller, 1984; Kelso, Holt, Kugler, & Turvey, 1980; Kugler, Kelso, & Turvey, 1980). Common features of the latter are that--like movement--they consist of very many subsystems. Unlike the theoretical approaches discussed above, however, where the emphasis is on detailed prescriptions for control, in synergetics, when certain conditions (so-called "controls") are scaled up even in very nonspecific ways, the system can develop new kinds of spatiotemporal patterns. The latter are maintained in a dynamic way by a continuous flux of energy (or matter) through the system (Haken, 1983). Although there is pattern formation in the nonequilibrium phenomena treated by synergetics, e.g., the hexagonal forms produced in the Bénard convection instability, the transition from incoherent to coherent light waves in the laser, the oscillating waves of the Belousov-Zhabotinsky chemical reaction, etc., there are strictly speaking no pattern generators. That is, the emphasis is on the lawful basis, including the necessary and sufficient conditions, for pattern formation to occur. The explanation is derived from first principles: it never takes the form of introducing a special mechanism--like a motor program--that contains or represents the pattern before it appears.

5. Phase Transitions in Biological Motion

There are already strong hints in the motor system's literature that a highly detailed prescription from higher neural centers is not necessary to produce either a stable spatiotemporal pattern (say among the legs of a locomoting animal) or an abrupt change in ordering among the legs, as in locomotory gait changes. An early indication comes from remarkable experiments by von Holst (1937/1973) on the centipede Lithobius. By amputating leg pairs until only three such pairs were left, von Holst transformed the centipede's gait (a pattern in which adjacent legs are about one-seventh out of phase) into that of a six-legged insect. Further, when all but two pairs of legs were left, the asymmetric gaits of the quadruped were exhibited. It is hard to imagine that the nervous system of the centipede possessed stored programs or pattern generators for these gaits in anticipation of its legs being amputated by an innovative experimenter. Rather, given a novel configuration, the system appears spontaneously to adopt those modes of locomotion that are dynamically stable. Synergetics attempts to predict exactly which new (or different) modes will evolve in complex systems particularly when the system undergoes qualitative macroscopic changes (Haken, 1983).

More direct evidence that rather diffuse inputs ("controls") can lead to highly ordered behavior comes from Russian studies on (decerebrate) locomoting cats (Shik, Severin, & Orlovskii, 1966). A steady increase in midbrain electrical stimulation was sufficient not only to induce changes in walking velocity, but also--at a critical stimulation level--to induce abrupt gait changes as well. Interestingly, unstable regions were also noted in which the cat vacillated between trotting and galloping.

A final clue suggesting that gait transitions belong to the class of nonequilibrium phase transitions comes from work on the energetics of horse locomotion. It is well known that animals use a restricted range of speeds (within a given gait) that corresponds to minimum energy expenditure. Hoyt and Taylor (1981), however, forced ponies to locomote away from these "equilibrium states" (see Figure 1) by increasing the speed of a treadmill on which the ponies walked. As shown in Figure 1, it becomes metabolically costly for the animal to maintain a given locomotory mode as velocity is scaled: for example, the walking mode becomes unstable, as it were, and "breaks" into a trotting mode (the next local minimum). Likewise, it is energetically expensive to maintain a trotting mode at slow velocities, a fact that appears to require switching into the walking mode (although no data on hysteresis are given). As in many other systems treated by synergetics, when a critical value is reached, the system bifurcates and a new (or different) spatiotemporal ordering emerges. Note that in Figure 1 these locomotory mode changes are not necessarily hard-wired or deterministic. Horses can trot at speeds at which they normally gallop, but it is metabolically costly to do so.

The notion that gait shifts correspond to instabilities that arise as the system is pushed away from equilibrium would be greatly enhanced if qualitatively similar phenomena were observed in other types of activities--perhaps even of a less stereotypical "innate" kind than locomotion. The remainder of this paper will be devoted to the elaboration of a phase transition that occurs in voluntary cyclical movements of the hands (Kelso, 1981, 1984). We will describe the phenomenon in Section 6 and illustrate briefly how it has been modeled using concepts of synergetics and the mathematical tools of nonlinear oscillator theory (Haken, Kelso, & Bunz,

1985). Finally, we will show that the phenomenon contains some of the principal features of other nonequilibrium phase transitions in nature. Interestingly, this synergetic account not only handles a variety of phenomena typically described by motor programs/CPG accounts, but also generates new predictions that have not come to light from either of these theories.

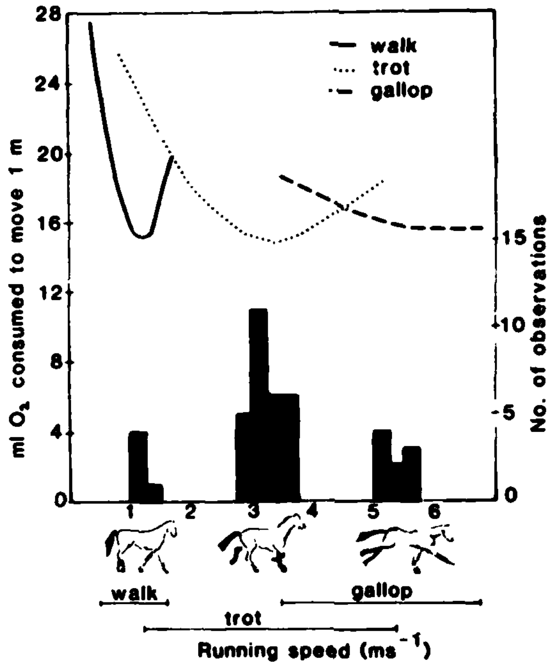


Figure 1. Oxygen consumption and preferred speed of walk, trot, and gallop of locomoting horses (see text for details). From Hoyt and Taylor (1981).

6. Nonequilibrium Phase Transitions in Bimanual Action

6.1 The Basic Phenomenon (Kelso, 1981, 1984; Kelso & Tuller, 1984)

In the bimanual experiments, a human subject was asked to cycle his/her fingers or hands at a preferred frequency using an out-of-phase, antisymmetrical motion. Under instructions to increase cycling rate, it was observed that at a critical frequency the movements shifted abruptly to an in-phase, symmetrical mode involving simultaneous activation of homologous muscle groups. When the transition frequency was expressed in units of preferred frequency, the resulting dimensionless ratio or critical value was constant for all subjects but one. This subject was not naive and purposely resisted the transition although with certain energetic consequences (see Kelso, 1984). A frictional resistance to movement lowered both preferred and transition frequencies, but did not change the critical ratio (~1.33). As an interesting aside, the ratio of transition speed to preferred speed for walk-trot and trot-gallop gait shifts, shown in Figure 1, also gives a value

~1.32. This dimensionless number (analogous, perhaps to a Reynolds' number in hydrodynamics) may provide a rough estimate of "distance from equilibrium."

In summary, the main features of the bimanual experiments are: a) the presence of only two stable phase (or "attractor") states between the hands (see also Haken et al., 1985; Kelso, 1979, for further evidence); b) an abrupt transition from one attractor state to the other at a critical, intrinsically defined frequency; c) beyond the transition, only one mode (the symmetrical one) is observed; and d) when the driving frequency is reduced, the system does not return to its initially prepared state, i.e., it remains in the basin of attraction for the symmetrical mode.

6.2 Modeling (Haken et al., 1985)

In complex systems it is clearly hopeless to try to investigate the motion of each microscopic degree of freedom. Rather the challenge is to identify and then lawfully relate singular macroscopic quantities to the interactions among very many sub-components. Close to instability points, it can be shown that the the behavior of the whole system is determined by one or a few order parameters (Haken, 1975). Such order parameters are not only created by the cooperation among the individual components of a complex system (e.g., by the interactions among atomic spins in a magnet), but in turn govern the behavior of those components (e.g., the magnetic field is an order parameter for a ferromagnet).

Identifying order parameters, even for physical and chemical systems, is not a trivial matter. Certain guidelines exist, however, that can be used for the selection of viable candidates. Two such selection criteria are: 1) the order parameter, by definition, changes much more slowly than the subsystems, i.e., its time constants are much longer than the time constants of the components; and 2) the order parameter's long term behavior changes qualitatively at the critical point.

In the case of our bimanual experiments and, we suspect, many other kinds of biological motion also, relative phase, ϕ, meets these criteria quite well (cf. Section 3.0). Using relative phase as an order parameter, Haken et al. (1985) modeled the bimanual data by specifying a potential function, V (corresponding to the layout of attractor states defined above), and showed how that function was deformed as a control parameter (corresponding to driving frequency) was changed. The choice of V--a superposition of two cosine functions--represented the simplest form that could describe the pattern of results. The series of potential fields generated for varying values of b/a (the ratio of the cosine coefficients) is shown in Figure 2. It can be seen that at a critical value, ω_c, the system jumps into a local minimum, i.e., there is a transition from the anti-phase mode ($\phi = \pm\pi$) into the symmetric, in-phase mode ($\phi = 0$). Moreover, the system stays in that minimum even where the driving frequency is reduced below ω_c, thus exhibiting hysteresis.

In an additional following analysis, Haken et al. (1985) used nonlinear oscillator theory to show how the model equations for the potential function could be derived from equations of motion for the two hands and a nonlinear coupling between them. Since the details are published we simply illustrate briefly some recent results of a consequent computer simulation (see also Haken et al., 1985, Figures 6 and 7).

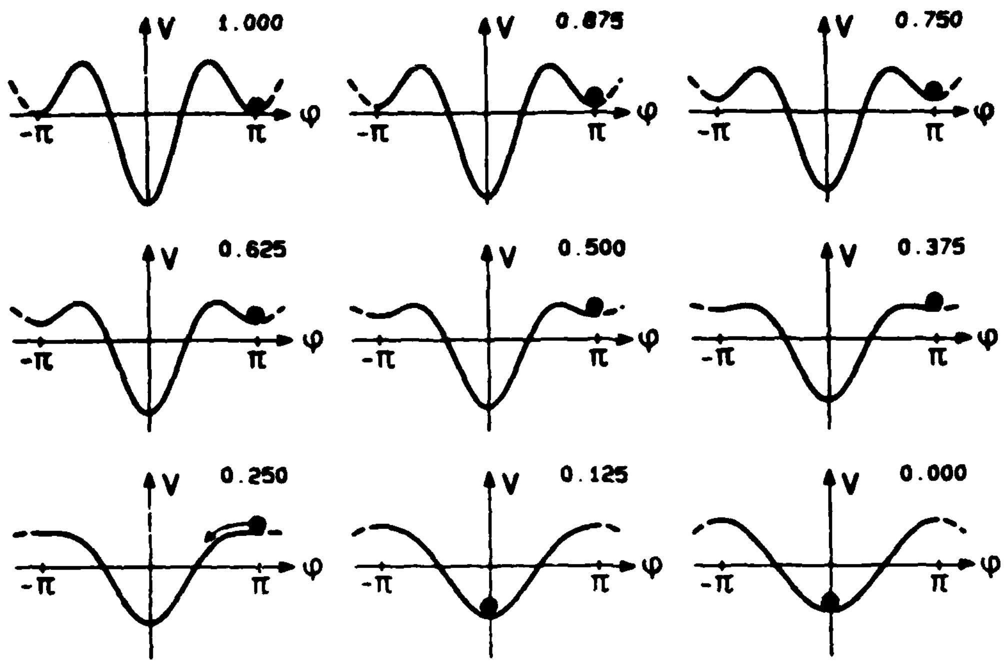


Figure 2. The potential V/a for the varying values of b/a. The numbers refer to the ratio b/a (from Haken et al., 1985).

In Figure 3, Lissajous portraits of the coupled oscillators are shown. The equations describing the motion are:

$$\ddot{x}_1 + (\dot{x}_1^2 - 1)\dot{x}_1 + kx_1 = \alpha(\dot{x}_1 - \dot{x}_2) + \beta(\dot{x}_1 - \dot{x}_2)(x_1 - x_2)^2 + F_{noise} \qquad (1)$$

$$\ddot{x}_2 + (\dot{x}_2^2 - 1)\dot{x}_2 + kx_2 = \alpha(\dot{x}_2 - \dot{x}_1) + \beta(\dot{x}_2 - \dot{x}_1)(x_2 - x_1)^2 + F_{noise} \qquad (2)$$

In (1) and (2) above the LHS corresponds to a Rayleigh-type, nonlinear oscillator (Equation 3.6 of Haken et al., 1985) the RHS is a Van der Pol coupling term plus some noise to simulate fluctuating forces (Equation 3.25 of (Haken et al., 1985). The only difference between the two simulations lies in the magnitude of fluctuations. Indeed, the transition shown in Figure 3(b) is remarkably like the behavior we observe typically (see e.g., Kelso & Tuller, 1984). Though we have not made a full study of the effects of initial conditions, coupling parameters, and fluctuations, our impression is that--given sufficient coupling strength--fluctuations play a major role.

Suffice it to note at this point that the model captures not only observed decreases in hand movement amplitudes as ω is increased, but also the abrupt change in qualitative behavior from antisymmetric to symmetric modes.

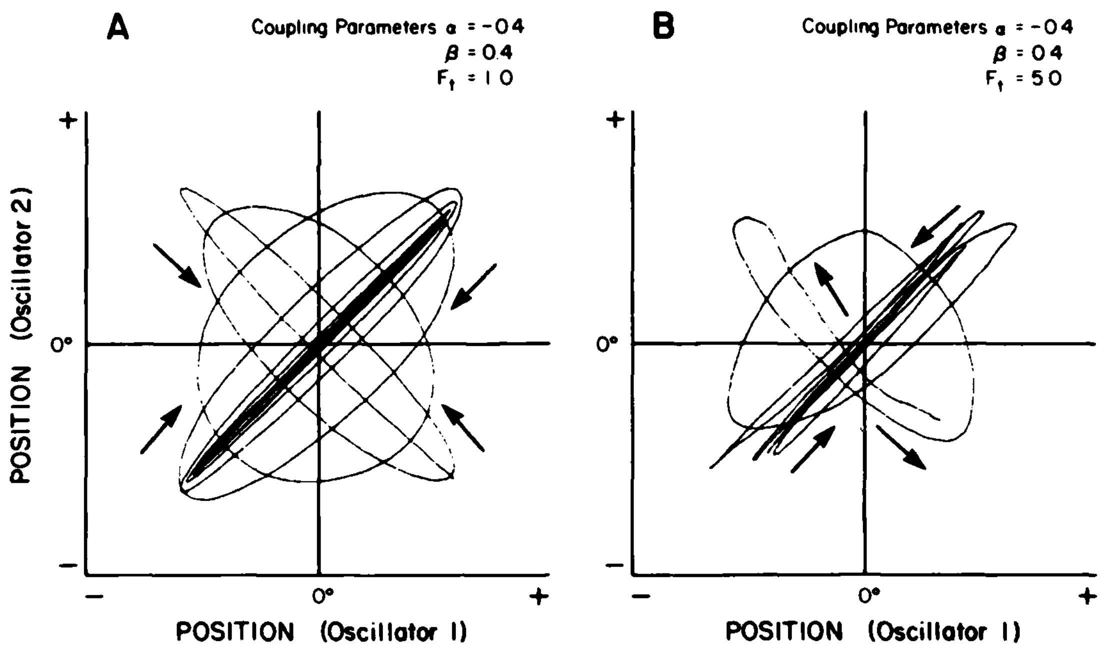


Figure 3. Lissajous portrait of behavior of two coupled Rayleigh oscillators (see text for details). Intrinsic frequency continuously scaled. Initial conditions of simulations: $x_1 = 25°$, $x_2 = -25°$, $\dot{x}_1 = \dot{x}_2 = 0$. A and B differ only in level of noise component. (We are grateful to Bruce Kay for performing the simulations).

6.3 Theoretical Underpinnings

If the bimanual phase transition constitutes a critical instability far from equilibrium, then certain specific predictions can be generated regarding the system's behavior near the transition. In particular, the hypothesized order parameter (relative phase) should exhibit at least two major properties: 1) critical slowing down as the transition is approached, i.e., the relaxation time of the order parameter to any perturbation should diverge at the transition. In general, the system exhibits a symmetry breaking instability, i.e., a constraint arises during the transition that restricts the future configuration of the system; and 2) enhanced fluctuations of the order parameter in space and time near the transition. The data presented next represent a preliminary attempt to explore the degree to which these theoretical predictions may or may not apply to phase transitions in hand movements.

6.4 New Experiments

We performed two kinds of experiment. In each, subjects were seated comfortably with pronated forearms, supported up to the metacarpal heads of the hand. The forearm was stabilized to restrict movement to the fingers alone. On each trial, the subject oscillated the index finger bilaterally in the transverse plane (i.e., abduction-adduction). Continuous finger displacement in the transverse and parasagittal (i.e., flexion-extension) planes was measured using a modified Selspot camera system. The electromyographic (EMG) activity of the right and left first dorsal interosseous (FDI) muscle was obtained with platinum fine-wire electrodes (see Figure 4). All data were recorded on a 12-channel FM-magnetic tape recorder for later off-line computer analysis.

Initially, subjects were instructed to move in one of two ways: oscillation of the right (R) and left (L) index fingers in either 1) the symmetrical mode or 2) the antisymmetrical mode, at their preferred rate. The frequency of oscillation was gradually increased to a maximum of approximately 3.5 Hz. In Experiment 1, the frequency of oscillation was increased every 2-3 s by asking the subject to increase his/her rate slightly. Thus, the rate of increase was not strictly controlled. In Experiment 2, the frequency of oscillation was systematically increased in 0.25 Hz steps every 4 s paced by a metronome. Data from trials in this experiment could therefore be averaged in time. Averages for Experiment 1 required alignment of trials by similar frequencies of oscillation. However, despite the lack of exact frequency equivalence, results from the two experiments are surprisingly consistent.

6.5 Order Parameter Behavior

6.5.1 Critical slowing down. The time series of one trial of finger oscillation, when the system is prepared initially in the antisymmetrical mode, is depicted in Figure 5a (note: the figure shows only a portion of the trial in the vicinity of the phase transition). Here, one can clearly see the transition to the symmetrical mode with an increase in the frequency of oscillation. In Figure 5b a point estimate of relative phase for the same sample record, based upon the peak displacement of the R and L fingers, is shown. A slow oscillation in phase, particularly before the transition, is evident. As the transition is approached, the frequency of this phase oscillation slows; the system takes longer and longer to return to its stationary state from a small deviation. This finding is a consistent feature of the experiments and is taken as preliminary evidence for the phenomenon of critical slowing down. Future work will calculate the relaxation time of the hypothesized order parameter explicitly using correlation techniques and perturbation experiments.

A continuous estimate of relative phase may be found in Figure 5c, based upon the continuous phase angle difference between each oscillator. Note that this estimate reveals some of the microscopic details of the phase fluctuations, while preserving the slow modulations in phase described above. A clear reduction in these fluctuations occurs following the transition. All remaining data on relative phase to be reported are based upon this continuous estimate.

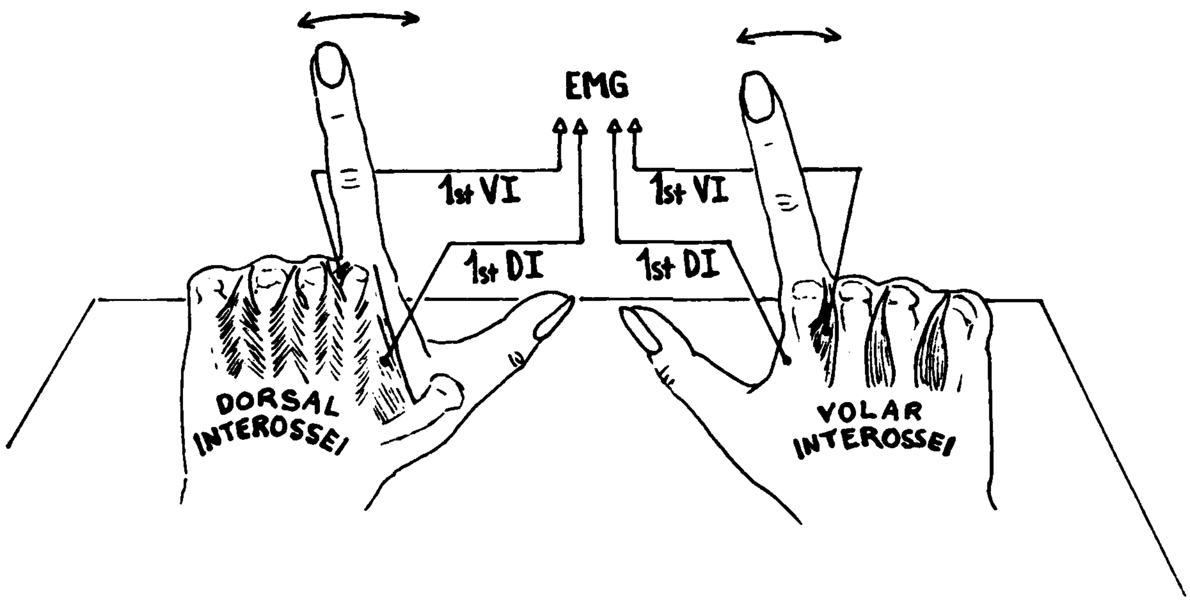


Figure 4. General experiment set-up for recording EMG. Support splints not shown (drawing by C. Carello).

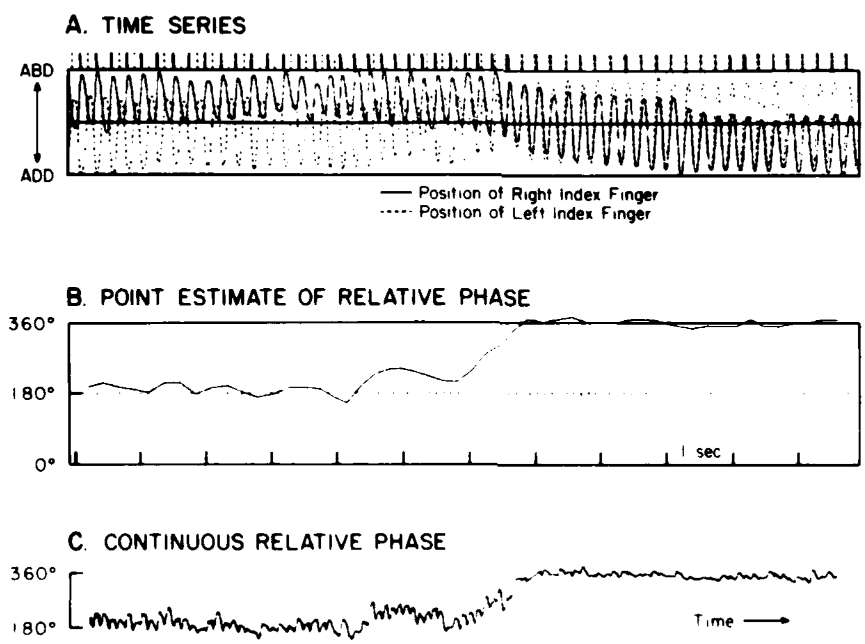


Figure 5. Time series (A) and relative phase (B & C) of R and L finger oscillation (see text for details).

6.5.2 Enhancement of fluctuations. An important feature of critical phenomena is the increase in variance of the order parameter near the phase transition. The system is said to become "soft" and thus unable to suppress critical fluctuations. The variance of the order parameter in the finger experiment is presented in Figure 6. The SD of continuous phase was calculated in the stable regime with the transient removed, i.e., over the last 3 s (= 600 data points) of oscillation at each frequency. Each point on the graph represents an average of 10 trials from Experiment 2. Mean phase is presented as well.

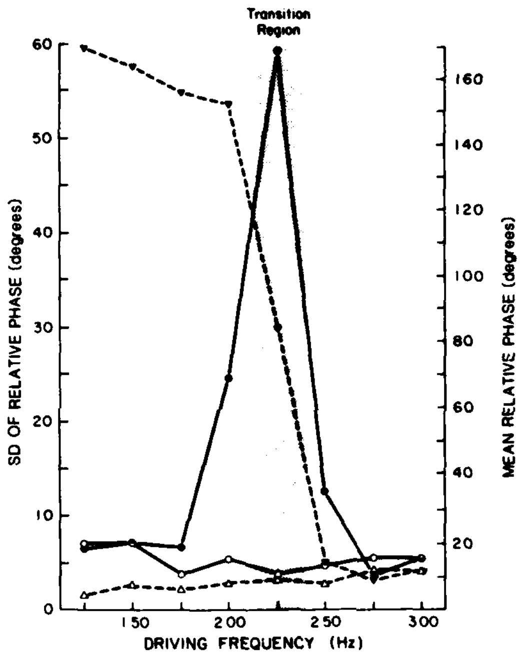


Figure 6. Mean (▼ AMS, Δ SMS) and standard deviation (● AMS, O SMS) of continuous relative phase at each driving frequency (n=10). AMS = antisymmetrical mode scaled. SMS = symmetrical mode scaled.[1]

Consideration of trials in which the system was initially prepared in the antisymmetrical mode reveals a clear increase in relative phase fluctuations as the transition is approached. The phase variance maximum at the transition is somewhat artifactual, since the phasing must change in order for a new mode to be exhibited. Note also that after the transition, the variance eventually stabilizes at a lower level (corresponding to the symmetrical mode) than before the transition. So-called control trials, in which the system is initially prepared in the symmetrical mode, exhibit no such increase in phase variance with increasing driving frequency. These findings are therefore consistent with theoretical predictions and the results of the nonlinear oscillator modeling shown earlier.

Order parameter dynamics can be further explored by examining the spectral content of relative phase. Each sample record of continuous relative phase was divided into eight segments corresponding to the increments in driving frequency. The power spectral density function (PSDF) of each segment was then determined by Fast Fourier Transform. Average PSDFs were obtained for trials in which subjects were initially prepared in the antisymmetrical mode, as well as those prepared in the symmetrical mode. The results are displayed in Figure 7. The DC component has been removed from each plot, since it represents the mean phase value, and overwhelms the other components, particularly in the anti-phase mode.

Figure 7a displays the average PSDF for trials initially prepared in the antisymmetrical mode. Note that as the driving frequency (ω) increases, a gradual increase in the frequency of the dominant spectral peak occurs. This increase appears to represent, in part, the influence of the driving frequency. Just prior to the transition, at 2.25 Hz, a dramatic increase occurs in the amplitude of the lowest frequency band, 0.8 Hz, along with the disappearance of higher frequency components. The stippled PSDF represents the transition region alone and reveals spectral broadening. With further increases in driving frequency the spectrum remains relatively broad and 0.8 Hz remains as a strong harmonic.

The average PSDF of trials initially prepared in the symmetrical mode is shown in Figure 7b. While higher spectral components are present as the driving frequency is increased, the 0.8 Hz component is always strong, even at low driving frequencies. Driving frequency appears to have relatively less effect on the PSDF of the symmetrical mode than that of the antisymmetrical mode. The dramatic increase in the amplitude of the 0.8 Hz component in the antisymmetrical mode just prior to the phase transition may represent the "swamping" of this mode's energy by that of the more stable symmetrical mode. That is, the longest lasting mode--symmetrical, in-phase--appears prominently before the transition itself. Though this interpretation is speculative at present, there does seem to be evidence that the antisymmetrical mode "feels" the driving frequency move strongly than its in-phase counterpart condition. In the language of synergetics, the order parameter is "slaving" its components less strongly in the former case than the latter.

6.6 Exploring the Neuromuscular Basis of the Transition

6.6.1 The η parameter. In order to determine the extent to which changes in EMG activity map onto those of the hypothesized order parameter already described, the parameter η was calculated. Figure 8a shows how this was done. R_0 and L_0 were obtained for each cycle of a sample record by determining the percent of total mean rectified EMG of one FDI that overlapped in time with that of the contralateral FDI. Note that η is thus a sample estimate of the total energy of motor unit activity within a time interval defined by the phase between the fingers. It therefore constitutes a way of observing how the "microscopic" quantities relate to the macroscopic phasing parameter. A plot of η vs. time (and increasing frequency) for one representative trial is provided in Figure 8b. The change in η maps quite nicely onto the change in the kinematic order parameter, as might well be expected. The η parameter change appears to occur more abruptly as compared to the change in relative kinematic phase, however.

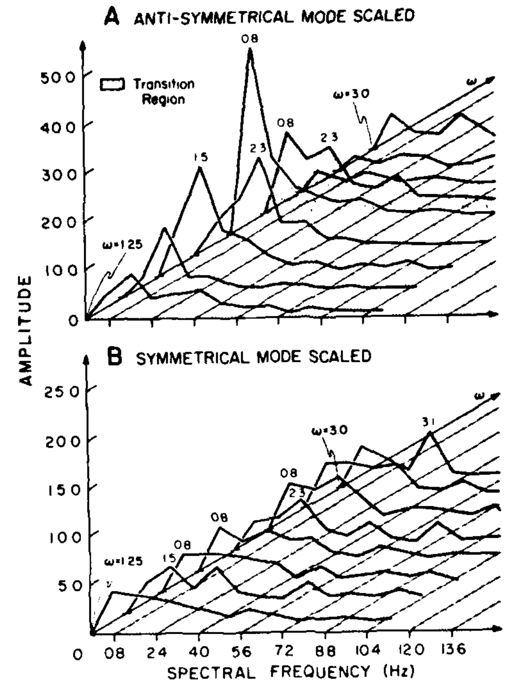


Figure 7. Average PSDF of continuous measure of relative phase computed at each driving frequency (ω) for trials prepared in A. antisymmetrical and B. symmetrical modes. (Note: ordinate of A and B has a different scale).

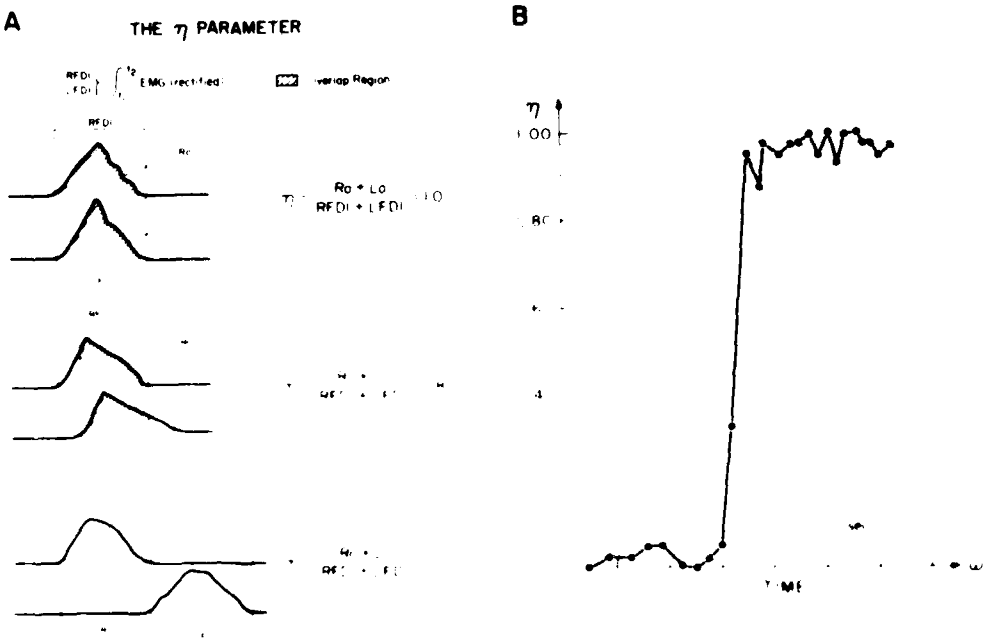


Figure 8. The η parameter. A. Method of calculation from mean rectified, integrated EMG. B. Plot of η vs. time (and increasing oscillation frequency ω) for one representative trial.

6.6.2 EMG autocorrelograms. One question concerns the nature of the neuromuscular reorganization underlying these phase transitions. In a preliminary attempt to examine this issue we looked at the autocorrelograms of mean rectified EMG for RFDI and LFDI, assuming they provide a measure of the temporal coherence of an individual muscle's activity. Two-second segments of sample records prior to, during, and immediately following the transition were analyzed. The calculation of each sample autocorrelogram was adjusted according to the oscillation frequency of the fingers so that the same number of peaks occurred in each function. The mean value of the peaks in each function and their coefficient of variation were calculated as measures of temporal coherence. Both measures yielded similar results.

The mean peak autocorrelation of seven trials (Experiment 1) is presented in Figure 9. The striking finding is the similarity between the coherence measures of the RFDI and LFDI before and after the transition, and their divergence at the transition. In the former two cases, even when the temporal coherence of one muscle is low, the contralateral FDI exhibits similar behavior. The correlation between the temporal coherence measures before and after the transitions was above 0.90. This presumably indicates a tight coupling of their activity patterns, even when operating antisymmetrically. By contrast, one muscle always becomes more or less coherent in the transition region. Here, correlation of the R and L coherence measure was low, negative and non-significant. Note also that the muscle showing the lowest coherence, and the direction of coherence change (compare with pre-transition measures) is never the same from trial to trial. Therefore, the underlying neurophysiologial mechanisms do not appear to be strictly deterministic as one might assume from a programming model of phase transitions.

6.7. Second Kinematic Phase Transition

As subjects move toward the upper extremes of oscillation frequency used in these experiments (~3.25-3.5 Hz), we have observed that a second instability occurs irrespective of the initial mode in which the subjects are prepared. In-phase modal behavior in the horizontal plane becomes unstable and gives way to a similar pattern in the vertical plane. A sample record of such an event is shown in Figure 10 in which the displacement of each finger in both horizontal and vertical planes is plotted versus time (and, therefore, increasing oscillation frequency). Motion frequently becomes rotary in nature before simultaneous flexion-extension occurs. Further analysis, using comparable procedures to those described above, is underway.

Note that in this situation there is an additional degree of freedom available for energy dissipation. Thus a new (or different) configuration among the oscillatory components can occur--an additional basin of attraction appears spontaneously. The basis for this second transition is not altogether clear and requires further exploration. It may be determined, in large part, biomechanically, linked to the relaxation times of the participating muscles (i.e., FDI and first palmar interosseous, FVI). As the frequency of oscillation increases, the relaxation times begin to exceed the 1/2 period of each cycle, resulting in maximum agonist-antagonist coactivity (Freund, 1983). Energy can no longer be dissipated through motion in the transverse plane. However, because the experiment left open an additional degree of freedom, parasagittal motion, the system adopts this new configuration, apparently in order to dissipate the increasing energy. Both the FPI and FDI have lever arms that provide contribution to finger flexion. The extent to which the long finger flexors and extensors are also facilitated cannot be determined by the present data.

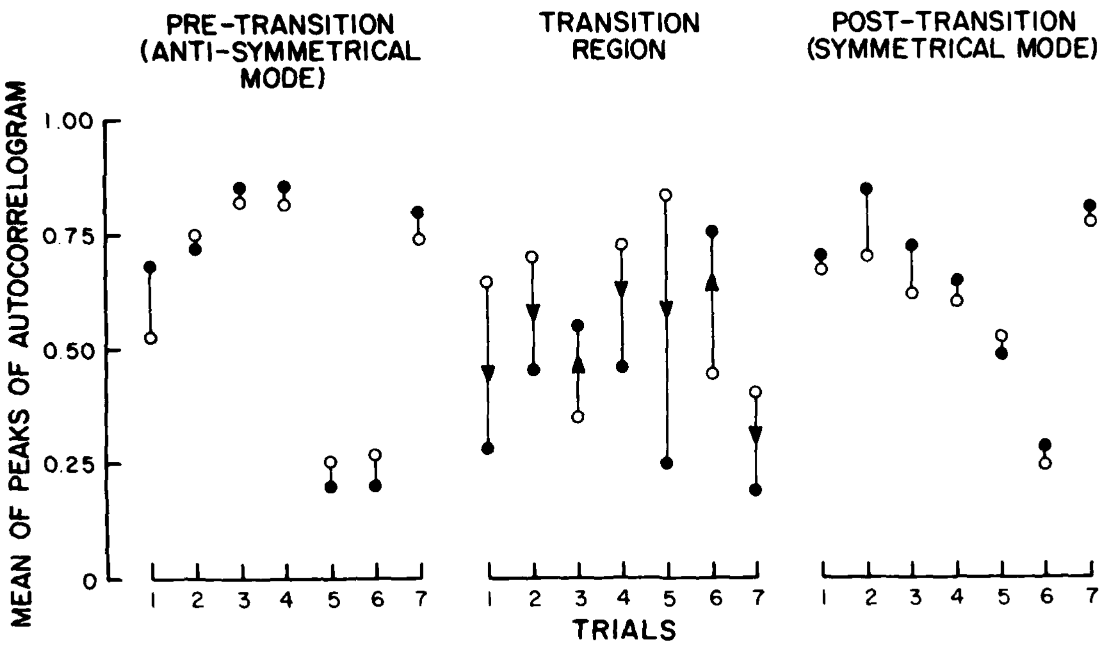


Figure 9. Measure of temporal coherence of right FDI (●) and left FDI (O) 2 s before, during, and 2 s after phase transition (see text for details).

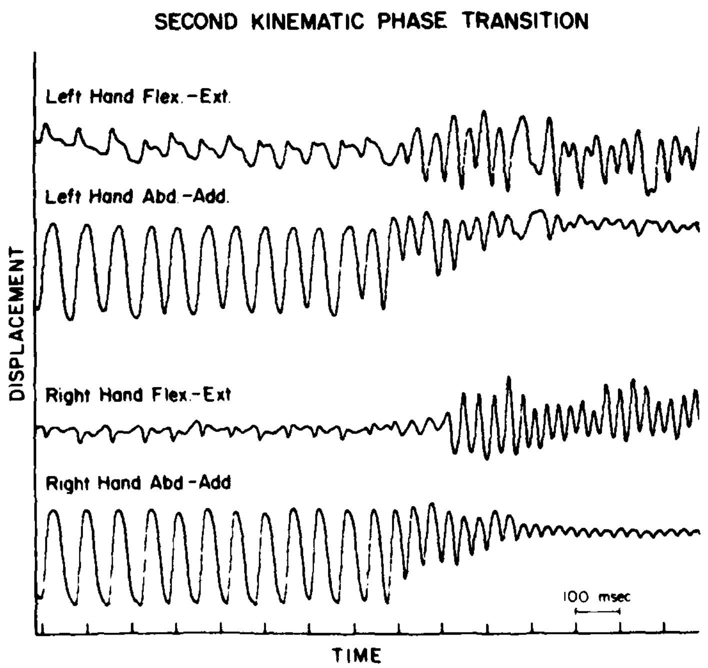


Figure 10. Time series of oscillation of R and L index finger in horizontal (abduction-adduction) and vertical (flexion-extension) planes for oscillation frequency above 3 Hz. See text for details.

7. Concluding Remarks

Neuroscience has not looked seriously to contemporary physical theory for ways to think about brain-behavior relationships. And, with few notable exceptions (this conference being one, see also Basar, Flohr, Haken, & Mandell, 1983), physics has made little contact with organic phenomena. Here we have shown, in a very preliminary fashion, how some of the tools and concepts of nonequilibrium phase transitions may offer insight into the emergence of space-time order at a macroscopic level. In our simple experiments we have begun to identify some of the main features of nonequilibrium transitions, including symmetry breaking, critical slowing down, and enhancement of fluctuations. Further work--both theoretical and experimental--will be necessary to converge on these and other characteristics, e.g., identification of the system's time scales and especially measurement of mode relaxation times using correlation functions and perturbation techniques, classification of the stochastic nature of fluctuations, exploring the system's sensitivity to parameter change, etc.

The central thrust here, of course, is to understand coordination in the multi-degree-of-freedom motions of animals and organisms. Even if we knew all the microscopic details about the system's components, we would still need a lawful description of how the components relate among themselves. An attraction of synergetics is that it deals with the formation of functional structures based on the cooperation among the system's many individual components. The theory achieves its full rigor when the system's behavior changes qualitatively, when newly emerging patterns are defined solely in terms of a few characteristic quantities, the so-called order parameters. A chief mechanism for the emergence of order lies in the competition between energy flowing into the operational components (i.e., a scaling influence) and the ability of those components to absorb the energy flow in their current configuration. As we have shown here (see e.g., Section 6.7) in the case of certain biological motions, higher bifurcations are possible if the system has available additional degrees of freedom, i.e., when a given configuration can no longer absorb the energy input. Moreover, fluctuations may permit the system's discovery of new modes or phasing structures.

If nature operates with ancient themes, as we suspect, then the same laws/strategies should appear at every level of description, and despite differences in material structure. Thus, the reductionism advocated here is not to any privileged scale of analysis, but rather to a minimum set of principles. The present treatment, preliminary though it is, may be just as pertinent to the mysteries of bacterial locomotion (see Janos, 1983) as it is to the coordinative patterns among the limbs and the abrupt transitions between them.

References

Basar, E., Flohr, H., Haken, H., & Mandell, A. J. (Eds.). (1983). Synergetics of the brain. Berlin: Springer-Verlag.

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon Press.

Bullock, T. H. (1976). In search of principles of neural integration. In J. D. Fentress (Ed.), Simple networks and behavior. Sunderland, MA: Sinauer Associates.

Bullock, T. H. (1980). Reassessment of neural connectivity and its specification. In H. M. Pinsker & W. D. Willis (Eds.), Information processing in the nervous system. New York: Raven Press.

Easton, T. A. (1972). On the normal use of reflexes. American Scientist, 60, 591.

Edelman, G. M., & Mountcastle, V. B. (Eds.) (1978). The mindful brain. Cambridge, MA: MIT Press.

Fowler, C. A. (1977). Timing control in speech production. Bloomington, IN: Indiana University Linguistics Club.

Freund, H.-J. (1983). Motor unit and muscle activity in voluntary motor control. Physiological Review, 63, 387.

Grillner, S. (1975). Locomotion in vertebrates. Physiological Reviews, 55, 247-304.

Grillner, S. (1985). Neurobiological bases of rhythmic motor acts in vertebrates. Science, 228, 143-149.

Haken, H. (1975). Cooperative phenomena in systems far from theral equilibrium and in nonphysical systems. Review of Modern Physics, 47, 67-121.

Haken, H. (1983). Syneretics: An introduction. Heidelberg: Springer Verlag.

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. Biological Cybernetics, 51, 347-356.

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. Nature, 292, 239-240.

Janos, L. (1983). Bacteria's motors work in forward, reverse and 'twiddle.' Smithsnian, 14, 127.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attentin and performance I. Hillsdale, NJ: LEA.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 15, R1000-R1004.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 49-70). New York: North-Holland.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., & Tuller, B. (1984). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience (pp. 321-356). New York: Plenum.

Kelso, J. A. S., & Tuller, B. (1985). Intrinsic time in speech production. Journal of the Acoustical Society of America, 77, S53.

Kelso, J. A. S., & Tuller, B. (in press). Intrinsic time in speech production: Theory, methodology and preliminary observations. In E. Keller & M. Gopnik (Eds.), Sensory and motor processes in language. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. Journal of Experimental Psychology: Human Percepion and Performance, 10, 812-832.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Loeb, G. E., & Marks, W. B. (1980). Epistemology and heuristics in neural network research. The Behavioral and Brain Sciences, 3, 556.

Mackay, W. A. (1980). The motor program: Back to the computer. Trends in Neuroscience, 3, 97-100.

Merzenich, M. M., & Kaas, J. H. (1984). Reorganization of mammalian somatosensory cortex following peripheral nerve injury. Trends in Neuroscience, 5, 434.

Procaccia, I., this volume.*

Selverston, A. I. (1980). Are central pattern generators understandable? The Behavioral and Brain Sciences, 3, 535-571.

Shapiro, D. C., & Schmidt, R. A. (1982). The schema theory: Recent evidence and developmental implications. In J. A. S. Kelso & J. Clark (Eds.), The development of movement control and coordination. New York: Wiley.

Shaw, R. E. (1981). Modeling chaotic systems. In H. Haken (Ed.), Order and chaos in nature. Heidelberg: Springer Verlag.

Shik, M. L., & Orlovskii, G. N. (1976). Neurophysiology of locomotor automatism. Physiological Reviews, 56, 465-501.

Shik, M. L., Severin, F. V., & Orlovskii, G. N. (1966). Control of walking and running by means of electrical stimulation of the midbrain. Biophysics, 11, 1011.

Sleigh, M. A., & Barlow, D. I. (1980). Metachronism and control of locomotion in animals with many propulsive structures. In H. Y. Elder & E. R. Trueman (Eds.), Aspects of animal locomotion. Cambridge: Cambridge University Press.

Tuller, B., & Kelso, J. A. S. (1984). The timing of articulatory gestures: Evidence for relational invariants. Journal of the Acoustical Society of America, 76, 1030-1036.

Tuller, B., Kelso, J. A. S., & Harris, K. S. (1982). Interarticulator phasing as an index of temporal regularity in speech. Journal of Experimental Psychology: Human Perception and Performance, 8, 460-472.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: LEA.

von Holst, E. (1973). On the nature of order in the central nervous system (1937). In The behavioral physiology of animals and man: The collected papers of Erich von Holst, Vol. 1 (Vol. ,). Coral Gables, FL: University of Miami Press.

Weisskopf, V. (1984). Frontiers and limits of science. Stiftung: Alexander von Humboldt.

Yamanishi, J., Kawato, M., & Suzuki, R. (1980). Two coupled oscillators as a model for coordinated finger tapping of both hands. Biological Cybernetics, 37, 219-225.

Footnote

[1]Note: Figure 6 was mistakenly labeled. It actually displays the point-estimate of relative phase. The continuous estimate exhibits the same behavior and may be obtained from the first author.

NATURALIZING THE CONTEXT FOR INTERPRETING SMA FUNCTION*

John P. Scholz,† M. T. Turvey,† and J. A. S. Kelso††

Clinical and experimental evidence presented in the target article supports the contention that the SMA plays an important role in the control and coordination of actions. The presence of the "alien hand sign" and difficulties initiating voluntary actions in patients with SMA damage appear to suggest a role in intentional processes. The evidence presented, however, does not support a model in which SMA serves to translate the intent to act into the "selection, linkage, initiation and anticipatory control of a set of 'pre-compiled' motor subroutines...." As the author notes, results of studies involving electrical stimulation, or lesions, of the SMA in subhuman primates are controversial. In addition, infarcts affecting SMA are rarely confined to this area alone, and diaschisis is undoubtedly an important factor in determining the behavioral manifestations of any brain lesion. It is also unclear how much can be concluded from studies of patients suffering intractable epilepsy in which the area of focal seizure activity, here the SMA, has been resected. Can one assume that other brain regions are functioning normally?

An understanding of the neural support for action will surely be fostered by behavioral studies of patients with documented lesions in restricted areas of the neuraxis. There is reason to question, however, the wisdom of any model of neural function that treats (1) a particular brain structure as functioning in relative isolation from the total system of which it is a part, and (2) a function as circumscribed by a particular brain structure. We concur with Schmitt (1978) that "...theories based on partial systems are subject to the component-systems dilemma that bedevils all attempts at biological generalization. Such theories fail to articulate and effectively deal with the essence of the problem, which is the distributive aspect that emerges from the complex interaction of functional units...in the brain" (p. 1). Nor are the roles of different brain regions necessarily distinct or fixed. Recent evidence from sensory mapping studies show, for example, that topographic cortical maps may move and change shape spontaneously, or in response to experience (Merzenich et al., 1984). What is important are the relational aspects among component processes participating in the generation

---

*Commentary on target article by Goldberg, G. Supplementary motor area structure and function: Review and hypothesis. The Behavioral and Brain Sciences, in press.

†Also Department of Psychology, University of Connecticut.
††Also Departments of Psychology and Biobehavioral Sciences, University of Connecticut.

Acknowledgment. This work was supported by ONR Contract N00014-83-C-0083 to Haskins Laboratories.

of an act (Fentress, 1984). As Bernstein (1967) argued, this will necessarily involve both traditionally conceived "motor" and "sensory" processes (although we agree with Gibson, 1966, and Reed, 1982, that this dichotomy is less than ideal).

Attempts to model CNS function with "machine" concepts may be misguided. In our view, notions such as motor programs, schemas and the like obscure rather than aid an understanding of the basis for the control and coordination of action (e.g., Kelso, 1981; Kugler, Kelso, & Turvey, 1980). A more principled attack on these issues follows the well-worn path of natural science. What are the physical strategies by which systems self-organize and by which cooperative states defined over very many microcomponents are assembled? And how might these strategies apply to the neuromuscular system in the production of voluntary acts? For example, primate movements exhibit discrete and rhythmic properties qualitatively similar to physical systems of quite different material structure, i.e., mass-spring systems (e.g., Bizzi, Polit, & Morasso, 1976; Fel'dman & Latash, 1982; Kelso & Holt, 1980). The coordinated unitary state of a pair of limbs, rhythmically oscillating at the same tempo, seems to be assembled through the conservations (of mass, energy and momentum) (Kugler & Turvey, in press). And transitions occurring from one gait to another in locomoting animals, as well as transitions found in bimanual coordination of humans, seem to obey principles similar to those determining phase transitions in nonanimate systems (Kelso, 1984). If movements are assembled and sustained through natural principles, then it is in the context of such principles that SMA function is to be understood. For example, how are these principles appropriately constrained? Does SMA function contribute nonholonomic constraints (i.e., constraints that temporarily restrict the system's trajectory from among the many possibilities). If so, how?

Similar qualms can be raised about equating the predictive control of behavior with internal models of possible linkages among events. In natural settings there is information available to specify how an animal must organize its neuromuscular system in order to achieve its goals (Gibson, 1979; Turvey & Kugler, 1984). Information relevant to the control of actions is available to and may be detected by a number of perceptual systems (e.g., auditory, haptic, visual, etc.) (Gibson, 1966, 1979). In the case of vision, information in the specificational sense is optical structure lawfully generated by the layout of surfaces and by movements relative to those surfaces. It contrasts with information in the injunctional/indicational sense (such as an instruction to push or pull), which is more nearly arbitrary than lawful. The author implies that the latter sense of information (1) underwrites intentional acts, and (2) constitutes the format for the space-time expectancies making up the predictive model. Neither implication seems warranted except, perhaps, in extreme cases. A stop sign provides information in the indicational sense. It informs the automobile driver that she or he must stop, but it does not tell the driver how to do so, i.e., when to begin braking, how hard to brake, etc. Fortunately, information specific to these control requirements is available to the driver in the optical flow field (Lee, 1976).

As intimated, information in the specificational sense is prospective. It informs an animal about the possibilities for action and about the outcomes of current action if present conditions persist. The importance of specificational information to the prospective control of actions has been shown in a number of recent studies involving different skilled actions and

different species (for reviews see Lee, 1980; Turvey & Kugler, 1984). Thus, the author's impression that vision functions retrospectively, primarily in a feedback mode, is surely off the mark. The upshot of the foregoing is that the author is evaluating SMA's role in intentional activity under a too restricted interpretation of prospective control.

Similarly, efforts to elucidate the role of neural processes in the generation of acts, and attempts to understand the deficits exhibited by patients with CNS damage, will be served better by natural, ecologically representative tasks (see also Kelso & Tuller, 1981, for similar arguments regarding apractic disturbances). For example, the author cites evidence from studies of Parkinsonian patients in support of his model. In general, these have involved visuomotor tracking tasks in which the visual target is a patch of light whose motions are arbitrarily constrained. While patients with Parkinsonism perform poorly in this task compared to normals, it is questionable to what extent the task touches upon the true functional deficit exhibited by these patients. It may be deceiving to draw conclusions from such artificial settings about how damaged brain regions function in normal situations where the informational basis for "predictive behavior" is largely law-based. Paradigms such as those developed, say, by Lee (for visuomotor coordination) and Nashner and colleagues (for postural-volitional relations; e.g., Nashner & McCollum, 1985) should not only illuminate SMA's functional significance in more natural tasks, but may also clarify its role in braiding the two kinds of information discussed herein.

References

Bernstein, N. A. (1967). _The coordination and regulation of movements._ New York: Pergamon Press.

Bizzi, E., Polit, A., & Morasso, P. (1976). Mechanisms underlying achievement of final head position. _Journal of Neurophysiology_, _39_, 435-444.

Fentress, J. C. (1984). The development of coordination. _Journal of Motor Behavior_, _16_, 99-134.

Fel'dman, A. G., & Latash, M. L. (1982). Afferent and efferent components of joint position sense: Interpretation of kinesthetic illusions. _Biological Cybernetics_, _42_, 205-214.

Gibson, J. J. (1966). _The senses considered as perceptual systems._ Boston: Houghton-Mifflin.

Gibson, J. J. (1979) _The ecological approach to visual perception._ Boston: Houghton-Mifflin.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), _Attention and performance_ (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. _American Journal of Physiology: Regulatory, Integrative, and Comparative_, _246_, R1000-R1004.

Kelso, J. A. S., & Holt, K. S. (1980). Exploring a vibratory systems analysis of human movement production. _Journal of Neurophysiology_, _43_, 1183-1196.

Kelso, J. A. S., & Tuller, B. (1981). Toward a theory of apractic syndromes. _Brain and Language_, _12_, 224-245.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), _Tutorials in motor behavior_ (pp. 3-47). Amsterdam: North-Holland.

Kugler, P. N., & Turvey, M. T. (in press). Information, natural law and the self-assembly of rhythmic movement. Hillsdale, NJ: Erlbaum.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Lee, D. N. (1980). Visuo-motor coordination in Space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North-Holland.

Merzenich, M. M., Nelson, R. J., Stryker, M. P., Cyander, M. S., Schoppmann, A., & Zook, J. M. (1984). Somatosensory cortical map changes following finger amputation in adult monkeys. Journal of Comparative Neurology, 224, 591-605.

Nashner, L., & McCollum, G. (1985). The organization of human postural movements: A formal basis and experimental synthesis. The Behavioral and Brain Sciences, 8, 135-172.

Reed, E. S. (1982). An outline of a theory of action systems. Journal of Motor Behavior, 14, 98-134.

Schmitt, F. O. (1978). Introduction. In G. M. Edelman & V. B. Mountcastle (Eds.), The mindful brain. Cambridge, MA: MIT Press.

Turvey, M. T., & Kugler, P. N. (1984). An ecological approach to perception and action. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. Amsterdam: North-Holland.

INFORMATION AND CONTROL: A MACROSCOPIC ANALYSIS OF PERCEPTION-ACTION COUPLING*

J. A. S. Kelso† and B. A. Kay††

1. Introduction

In this chapter we address problems pertaining to the control of action--problems that, fundamentally, rest with understanding how perception and production are linked in biological activities. There have been a number of quite recent treatments, both behavioral and physiological, of motor control of simple limb movements performed in relatively uncomplicated environments. Rather than review that material again (see, e.g., Keele, 1981; Kelso, 1982a; Schmidt, 1982, for largely behavioral treatments; and, e.g., Houk & Rymer, 1981; Stein, 1982, for a largely neurophysiological-engineering analysis), we shall try to expand the horizons of "control" a bit in this chapter--a larger sweep of the brush, as it were (see also Reed, 1982). To a certain extent, we shall consider goal-directed activities like reaching for a cup, driving a car, climbing stairs--activities that involve very large numbers of degrees of freedom on both the motor and perceptual side of things. Thus, on the performance side were one to count, say, the number of neurons, neuronal connections, and muscle fibers involved (even in so-called simple actions like moving a finger), the result would be a large number. Likewise, on the perception side the light rays to the eye, the retinal mosaic, and the neural processing structures involved amass into a problem of huge dimensionality. Yet somehow--in spite of the large dimensionality on both sides of the coin (or perhaps because of it)--control is possible. Somehow, this high dimensionality gets compressed, as it were, into lower dimensional control. How this is realized, of course, is the challenge faced, not only by students of perception and action, but in other realms of science as well.

In this chapter we shall have this challenge in focus as we (1) present what an understanding of control in the larger context of perception-action systems might entail; (2) show how an approach based in dynamical systems theory can, on the action side, offer useful ways to describe the behavior of multi-degree of freedom systems; and (3) using concepts developed in (2) along with recent empirical analyses of visually guided actions, try to reveal the nature of the linkage between perceiving and acting. Questions such as: What <u>kind</u> of information is used to regulate action? <u>When</u> and <u>where</u> in a given

---

*In H. Heuer & A. F. Sanders (Eds.), Tutorials in perception and action. Amsterdam: North-Holland, in press.
†Also Departments of Psychology and Biobehavioral Sciences, University of Connecticut.
††Also Department of Psychology, University of Connecticut.
Acknowledgment. Much of the work referred to herein was supported by NIH Grant NS-13617, Biomedical Research Support Grant RR-05596, and U.S. Office of Naval Research N0014-83-C-0083. We appreciate very helpful comments by Herbert Heuer, Elliot Saltzman, and Michael Turvey on an earlier draft of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]

action is such information used, and how is it used? will receive our primary attention. We argue, as have our colleagues (Fitch & Turvey, 1978; Kugler, Kelso, & Turvey, 1980; Kugler & Turvey, in press; Saltzman & Kelso, 1983a; Solomon, Carello, & Turvey, in press) that by appropriate macroscopic descriptions of perceptual and motor parameters, the potentially complex, high dimensional control problem seen at the level of the microscopic degrees of freedom can be simplified.

Before proceeding we should mention that in making these moves, we stand on the shoulders of giants. On the perception side, Gibson (1961, 1966, 1979) developed the idea of the optical flow field as a relevant macroscopic description of the light to an eye (any eye) that is specific to the layout of surfaces and the activity of a moving point of observation. On the action side, Bernstein--at least in his later work (1967; Whiting, 1984)--pursued a macroscopic analysis of movement in terms of the essential and nonessential parameters governing large ensembles of neuromuscular elements, namely, those parameters that remain invariant during the course of an activity and those that do not. In each case, as we shall see, singular macroscopic quantities emerge that play a key role in the control of activity. But first, let us turn briefly to the meaning of control--both in its conventional form, as something that is imposed on a system by external means--and in the way we would like to view it, as arising intrinsically from the dynamics of the perception-action system itself.

2. Control

The concepts of regulation and control have played a central role in efforts to understand how the many neuromuscular degrees of freedom are harnessed to produce coherent behavior. In a cybernetic system, regulators and controllers serve closely related yet quite distinct functions. On the one hand, given a desired state of affairs in such a system, and a source of variability that can perturb the system away from that state, a regulator maintains that state within acceptable tolerance limits. For example, a thermostat regulates an oven's most important state variable, temperature, in the face of heat fluxes perturbing that temperature. On the other hand, control presupposes the existence of regulation capabilities in a system: the controller sets the particular values that the regulator tries to maintain. As a prosaic example, a chef controls a thermostat on an oven, to cook a meal slowly at a low temperature or more quickly at a higher temperature. Control function is most often provided by a logical separation between the controlling device and the controlled system (i.e., the plant dynamics). Hence, it is not appropriate to consider the controller to be a part of the system in the same sense as a regulator: whereas a regulator must be sensitive to apposite aspects of the system's dynamics in order to function at all, the specification of control algorithms is in principle arbitrary with respect to those dynamics (see Tomovic, 1978, for informed discussion of the plant-controller problem). Thus, the controller is extrinsic to the system and prescribes the system's behavior.

In the motor systems' literature, we see this view of control quite clearly expressed, for example, in Stein's (1982) article in The Behavioral and Brain Sciences on "What muscle variable(s) does the nervous system control?" in limb movements. For Stein and others (see Commentaries, ibid) the skeletomuscular apparatus is the system being controlled, and it is assumed that the nervous system is the controlling device. Control proceeds prescriptively, according to executive command programs, for example. We have

argued (e.g., Kelso & Saltzman, 1982; Kugler et al., 1980) that such a strategy offers little explanatory power, since it attributes the coherence and adaptability of coordinated movements to the coherent actions of an external controller, actions which themselves are not explained. Thus, control in this classical engineering sense is an example of allonomy, literally, "external law." Its complement is autonomy or "self-law" (Varela, 1979). Successful biological systems are autonomous in that no external controllers are necessary for their survival. Energy flows figure significantly in the survival of any organism (Morowitz, 1968), and as Yates, Marsh, and Iberall (1972) argue, in order to obtain efficient operation, a controller must be coupled to the system being controlled via an appropriate match or scaling between the energy flows of controllers and controlled systems. This criterion of energy flow commensurability applies to any control situation in which systems dissipate significant amounts of energy, a condition satisfied for biological motions. The criterion is clearly <u>not</u> met by the cybernetic theory of control and regulation, in which low energy signals (e.g., in microprocessor circuits) prescribe the large energy flows for the controlled systems (e.g., in torque motors for industrial robot arms). However, autonomous control, in which control resides "inside" the system as a natural consequence of its self-organization, does afford the possibility of satisfying this energy commensurability criterion.

Allonomic control theories imply an extrinsic view of control precisely because of the way they compartmentalize systems. For example, the perceptual and motor "apparatuses" are treated as fundamentally distinct components of a larger system (an organism), and organisms and their environments are also treated separately. Decompositions of this kind, though the trademark of analytic reductionism, can carry serious consequences for measurement and understanding (see Rosen, 1978). The problem is that such decomposition obscures the nature of the overall system's dynamics: an analysis of the system's parts may not lead to an understanding of the behavior of the system as a whole. Furthermore, the observables chosen to describe the parts may have nothing to do with those that are appropriate for the description of the system in toto. We are not repeating here the well-known adage that the whole is greater than the sum of the parts. Rather, we want to emphasize that in open, complex, multi-degree of freedom systems, novel properties, which cannot be known or predicted from knowledge of component processes, emerge at more global levels. Thus, not only do we have <u>more</u> of something as complexity increases, but that "more" is different (Anderson, 1972). This is an inevitable consequence of <u>broken symmetry</u>: systems with large numbers of microscopic degrees of freedom may undergo sharp, discontinuous transitions leaving behind usually few, qualitatively different modes of behavior. Such systems are subject to constraints that arise during the transitions, and thus cannot assume all those configurations that were possible before symmetry breaking. We shall return to this theme later because it affords a way of intuiting how the degrees of freedom of perception-action systems can be "compressed" as it were, so that coordination may be defined over a smaller number of variables.

One major consequence of viewing control as autonomous and self-organized is that the definition and role of <u>information</u> is drastically changed. In conventional control theory, information is arbitrary with respect to the activities that it serves. More generally, neither environmental events nor the perceiver's own movements are assumed to structure perceptually relevant energy distributions in ways that are intrinsically meaningful to the organism. Rather, information must be interpreted and disambiguated. An

autonomous view of control, however, mandates that information be: a) unique and specific to the facts about which it informs, b) meaningful to the control requirements of the activity (i.e., it carries its own "semantics" as it were), and c) scaled to the system's physical dimensions and behavioral repertoire (see Kugler, Kelso, & Turvey, 1982). In a deep sense, information for a self-organizing, autonomous theory of control is "in-formation," that is, the formation of structure in the system as a whole (Varela, 1979). In the present context, of course, the system is the perception-action system.

But how can we understand information as viewed within such a framework? How can these formulations be grounded in experimental analyses? To proceed further, we must make one additional, yet perhaps crucial, distinction--namely between a view of information as indicational/injunctional and a view of information as specificational.

3. Information

Information theory is still a powerful tool in many branches of science where it is used to obtain a measure for the amount of information contained in a system. It has had its application in the motor skills field as well, particularly through the stimulus of the late Paul Fitts (e.g., Fitts, 1954). Here is not the place to discuss the details of this theory except to make a few points. First, the formalisms derived from information theory (e.g., $I = k*\log(R_0)$, where I is the information metric, R_0 is the number of equiprobable events and k is an arbitrary constant) refer to the scarcity of an event; "information" is thus a measure of ignorance about a system (Ashby, 1956). Second, the events dealt with in information theory are symbolic, not dynamic, events. Even in physics, and certainly in other fields like biology and psychology, "information" takes the form of a set of symbolic elements organized by a grammar. The role that such symbolic structures play can be termed injunctional/indicational (see Reed, 1981; also, Kugler et al., 1982; Turvey & Kugler, 1984). On the one hand states may be indicated symbolically and, on the other, states can be commanded. In contemporary theories of motor control, for example, the motor program tells the muscles when to turn on, how much, and when to turn off. Emphasis here is clearly on the injunctional mode of description with little or no attention given to the rate-dependent, dynamical processes that are prescribed to or directed by the injunctional mode. Further, the symbolic or indicational mode of description greatly underestimates (to the point of ignoring) the information actually required to perform an activity. As Turvey and Kugler (1984) note, a stop sign indicates to a driver that the car should be stopped, but provides no information about how to stop the car, that is, how, where, and by how much to decelerate, apply the brakes, etc.

But as suggested above and as repeatedly emphasized in the writings of Pattee (e.g., Pattee, 1972, 1973, 1977), complex systems (the focus here) are to be fundamentally understood in terms of two complementary modes of description--the discrete, symbolic rate-independent mode and the continuous, dynamical, rate-dependent mode where the flow of time is included. In spite of the dualism implied by complementarity, the significance of Pattee's analysis for students of perception and action (see Kugler et al., 1982) is his emphasis on dynamical processes. That is, information in the symbolic sense plays a minimum-role; it acts as a constraint on dynamics but does not explicitly control them. Thus, although both modes of description are crucial to Pattee, the dynamical mode should be exploited to the fullest. Paraphrasing Emerson, hitch your wagon to a star--and see the chores done by

the Gods themselves (quoted by Greene, 1982, in the context of arm movement control).

As we have noted above and elsewhere (e.g., Kelso, Holt, Rubin, & Kugler, 1981) most of the theoretical effort in the field of movement science has stressed the symbolic, indicational mode. The contribution of dynamical processes is given a fairly limited treatment. For example, there have been many proposals for the "contents" of the motor program (see Kelso, 1981, for a critical review of putative candidates). Little attention has been paid to the processes by which these "contents" interface to the large-scale muscular machinery that carries out their instructions. More important, such theorizing lacks a rationale for how it is and by what means the particular contents of the program are created. What is missing is an account of the program that is priviledged with respect to the dynamics that it directs. The origins of the program's code must, it seems, be lawfully derived from dynamics (see Kugler et al., 1982; Turvey & Kugler, 1984). In summary, what we are saying amounts to this: 1) Information in the conventional, symbolic sense is not sufficient to control ongoing action; 2) Ergo, information in a nonsymbolic sense must play a significant role; 3) Such information is dynamical in the sense that it is unique and specific to the dynamics of activities themselves. That is, information is implicit in the dynamics, not imposed upon it as a sequence of symbol strings from the outside. In the following sections we provide a short tutorial of what is meant by dynamics, list some of the advantages of dynamic description, and provide some specific examples of its use in the movement field.

4. Introduction to Nonlinear Dynamics

Nonlinear (qualitative) dynamics is fundamentally concerned with the appropriate description for forms of motion in complex, multidegree of freedom systems. These forms of motion are specified, roughly, by the qualitative shapes observed in phase portraits of a system's behavior. The phase portrait constitutes the totality of all possible phase plane trajectories generated by a particular dynamical system under a particular parameterization. Phase plane trajectories have been used to varying degrees by engineers over the years, though their full significance is just being realized--at least in the West (see Abraham & Shaw, 1982, for a brief historical treatment). On the other hand, many developments in nonlinear dynamics have been pioneered by Russian workers (e.g., Andronov & Chaikin, 1949; Minorksy, 1962).

A phase plane trajectory is generated by plotting the position (x) of an articulator (say the end of a finger, the tip of the tongue, etc.) against its instantaneous velocity ($\dot{x}$). These quantities act as coordinates that describe the ongoing motion of the articulator in two-dimensional space; for a (deterministic, classical mechanical) system composed of one macroscopic degree of freedom, these two variables represent the state of the system at any point in time. As time varies, the point $P(x,\dot{x})$ moves along a certain path or trajectory on the phase plane. For different initial conditions (such as a given starting position) and parameter values (such as a given level of articulator stiffness) the motion will describe different phase paths. For a given system and set of parameter values, the form of the phase portrait (the ensemble of all the trajectories arising from all possible initial conditions) is specified by the relations among underlying dynamic parameters (for examples, see below). Such patterned forms or topologies can be categorized as low-dimensional attractors even though the system they describe is high dimensional.

This brings us to an important point: one reason, it seems, why dynamics has been of little interest to motor behavior theorists is that it has been conceived as local and concrete, pure biomechanics as it were. This bias is misplaced: dynamics, by definition, constitutes the simplest and most abstract description of the motion of a system (Maxwell, 1877/1952, p. 1). There is no logical reason why dynamics, although rate-dependent and nonsymbolic, cannot be abstract. Quite to the contrary, as any cursory perusal of the field of dynamical systems will reveal (e.g., Guckenheimer & Holmes, 1983; Haken, 1983; Rasband, 1983). Indeed, as many researchers are now discovering, complex systems composed of very different materials can share the same underlying dynamic structure (for many examples in physics, chemistry, and biology, see Haken, 1975, 1977; in movement science, see Kelso & Tuller, 1984a, 1984b).

An example of the dynamical approach in the field of motor systems was Fel'dman's (1966) insight that, in certain types of tasks, the motor apparatus behaves in a qualitatively similar way to a simple physical system, a mass-spring. Although a system of neuromuscular components differs greatly from a system of masses and springs, they can be shown to share the same abstract functional organization, that is, an equivalent dynamic, that of Hooke's law relating stresses and strains. As Rosen (1970) remarks, there is nothing unscientific or speculative about the dynamic approach, any more than, say, the hard sphere model for describing the behavior of gases, regardless of each gas's individual molecular structure. Indeed, if one's primary focus is function and behavior, then it is the search for appropriate dynamical descriptions of system behavior that takes precedence over any particular material embodiment. Such a strategy has played a major role in the development of science. Prigogine and Stengers (1984), for example, propose that Fourier's law, a mathematical description of the propagation of heat in materials (proposed in 1811), was the start of "a science of complexity" (p. 104). This simple law, which states that heat flow is proportional to the gradient of temperature, applies to all matter regardless of its state--solid, liquid, or gas. Also, the chemical composition of the substances to which it applies is immaterial; although each substance has its own proportionality coefficient, the same law holds nevertheless. Here again we see that in spite of a great deal of diversity at a molecular level, the macroscopic behavior is described by a single law, with particular variants resulting from changes in only a single parameter. The framework of nonlinear dynamics follows this macroscopic, law-based orientation to microscopic diversity. It offers a way of characterizing regularities in action problems in terms of relatively abstract, functionally specified control schemes.

5. A Brief Survey of Nonlinear Dynamics Applied to Movement Control

5.1 Generative Properties and Low-dimensional Control--Point Attractors

Attractors represent the asymptotic behavior of a whole family of system trajectories. As a simple example, referred to briefly above, a damped mass-spring system with only a single degree of freedom can have many trajectories depending on its initial conditions and its parameter values. For example, the linear mass-spring system

$$m\ddot{x} + b\dot{x} + kx = 0 \qquad (1)$$

may simply oscillate without being damped out (if the linear damping term, b, equals zero), or be underdamped, overdamped, or critically damped, depending

on the mass (m), the damping (b), and stiffness (k) parameter values (for actual examples of discrete movements displaying these types of behavior, see Kelso & Holt, 1980). For b greater than zero (corresponding to a real system having some frictional component), such a system is called a point attractor, a generic dynamical category that reflects the fact that all trajectories converge to an asymptotic, static equilibrium state (see Figure 1a). Such systems exhibit the property of equifinality--the tendency to achieve an equilibrium state regardless of initial conditions. Importantly, however, a multidegree of freedom system whose trajectories converge to a single rest position can also be described as a point attractor. One can imagine, for example, the high dimensionality involved in a simple finger movement, were one to include the neurons, muscles, and their interconnections, yet the resultant behavior would be described as a low-dimensional point attractor. Thus, point attractors also provide low-dimensional descriptions of the asymptotic patterns produced by potentially high-dimensional systems.

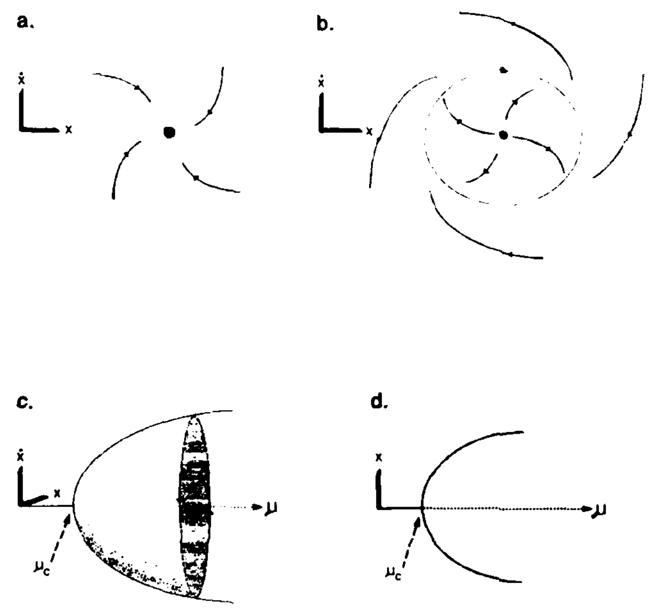


Figure 1. Phase plane portraits for a) a point attractor and b) a limit cycle oscillator. Bifurcation diagram of the c) Hopf and d) pitchfork bifurcations: as the parameter μ is increased, behavior shifts from a point attractor regime to a periodic regime, in two dimensions for the Hopf and one dimension for the pitchfork bifurcation.

Saltzman and Kelso (1983b) have recently shown how a point attractor dynamical regime defined at a task level can control the behavior of a multidegree-of-freedom system in such activities as reaching, cup-to-mouth tasks, and postural stability (see also Saltzman & Kelso, in press). This demonstration seems significant given criticisms that the mass-spring model (so-called "end-point control," Bizzi, Chapple, & Hogan, 1982) for single-joint motions is inadequate for motions involving two or more joints

(e.g., the arm and shoulder). The latter display (roughly) straight line trajectories of the hand (e.g., Bizzi et al., 1982; Morasso, 1981). However, though point attractor dynamics defined for each joint could generate the final target configuration, they would also result in a curved rather than quasi-straight line trajectory of the hand.

Part of the problem here may be the narrow definition of the mass-spring model. Some (a little naively, we note with 20/20 hindsight) have restricted the model to single, discrete movements in which muscles are represented by a pair of springs acting across a hinge in the agonist-antagonist configuration. The final equilibrium point is established by selecting the length-tension characteristics of opposing muscles (e.g., Bizzi, Polit, & Morasso, 1976; Cooke, 1980; Kelso, 1977; Schmidt & McGown, 1980). This view, at best, may work for deafferented muscle but, as pointed out by Fel'dman and Latash (1982) it is inadequate for muscles operating in natural conditions. Clearly, the parallel between a single muscle and a spring should not be taken too literally. The mass-spring model--as intimated above--is better viewed as an account of equifinality, a property shared by mass-springs and a complex, multivariable system's ability to generate targeting behavior (see Kelso, Holt, Kugler, & Turvey, 1980). By adopting this approach and specifying point attractor dynamics in task space, Saltzman and Kelso (1983b) show how sets of dynamic parameters, which are constant at the task level, can be used to define changing patterns of dynamic parameters at the articulator level (e.g., joint stiffness, dampings, rest angles). Thus, via this strategy a low-dimensional control scheme is realized that possesses generative properties. Once the relations among dynamic parameters are set up according to particular task demands, a wide variety of trajectories can be generated. Moreover, this rich set of trajectories emerges from an underlying task dynamic that does not contain detailed, step-by-step trajectory plans (e.g., Hollerbach, 1982) of any kind.

Thus, just as early work on single discrete motions showed that variables like duration and velocity did not need to be conceived as contents in the motor program, but were rather consequences of a simple, point attractor (mass-spring) dynamical system (e.g., Fitch & Turvey, 1978; Fowler, Rubin, Remez, & Turvey, 1980; Kelso, 1977; Kelso & Holt, 1980; Kelso et al., 1980; Schmidt & McGown, 1980), so this recent extension of dynamics by Saltzman and Kelso (1983b) demonstrates how program candidates for two-joint motions (such as trajectory) can arise from an appropriately specified dynamical regime. A very similar analysis holds for tasks involving multi-degree of freedom interlimb coordination (Kelso, Putnam, & Goodman, 1983; Kelso, Southard, & Goodman, 1979).

5.2 Generative Properties and Low-dimensional Control--Periodic Attractors

The theme that kinematic diversity can arise from an underlying "simple" dynamic control structure can be readily extended to rhythmical movements. Several years ago, we showed that bimanual, cyclical movements of the hands possess behaviors that are realizable by coupled nonlinear limit cycle oscillators (Kelso et al., 1981). Of course, a variety of rhythmical behaviors, such as locomotion in both vertebrates (e.g., Miller & Scott, 1977; Patla, Calvert, & Stein, in press; Willis, 1980, for reviews) and invertebrates (e.g., Cohen, Holmes, & Rand, 1982) can and have been modeled in similar ways--far more explicitly in fact than in the Kelso et al. (1981) paper (but see Haken, Kelso, & Bunz, 1985).

The limit cycle oscillator is called a periodic attractor in the dynamics literature because it displays orbital stability. Like a point attractor, all trajectories converge to a single limit set, in this case, a single cyclic orbit on the phase plane $(x, \dot{x})$, the limit cycle (see Figure 1b). "Equifinality" for a limit cycle is caused by a nonlinearity in the damping term (sometimes called the escapement). If the system's initial conditions are outside the limit cycle, the trajectories decay until they reach the limit cycle. Energy is dissipated until a balance between kinetic and potential energy occurs. Likewise, if the initial conditions are inside the limit cycle, trajectories grow or spiral out to the attractor (see Jordan & Smith, 1977; Minorsky, 1962). Mathematically, there are many kinds of equations describing stable periodic motion, most typically in differential form like equation (1). However, they are all topologically the same, that is, they all exhibit orbital stability, because the structure of the equations (in terms of the internal relations among parameters) is identical, although the parameters values themselves may change. It is the feature of topological invariance[1] that allows for the classification of dynamical systems into generic categories (Abraham & Shaw, 1982), and that perhaps affords a classification of movement tasks as well (for examples see Kelso & Tuller, 1984a; Saltzman & Kelso, 1983b).

In some cases a single parameter in a dynamic control structure can regulate the space-time behavior of the system. In recent work at Haskins Laboratories, we have investigated how spatiotemporal changes occur in single and bimanual cyclical movements in response to an externally required change in frequency. We wanted to try to understand a very basic question (but for which little information exists in the literature, see Freund, 1983): How do space (in terms of movement amplitude) and time (in terms of movement duration) covary as the task requires the hands to move faster? Subjects performed cyclical movements in response to a metronome whose frequency was manipulated (in 1 Hz steps) between 1 and 6 Hz. Subjects grasped handles with one or both hands--the forearms were stabilized and the task required movement around the wrist joint(s) in the horizontal plane. Transducers situated above the axes of rotation of the joints provided ongoing measures of angular displacement over time. The data on four subjects tested on two separate occasions revealed a reciprocal relationship between cycling frequency and amplitude for both single and bimanual movements (Kay, Kelso, Saltzman, & Schöner, submitted). Using a nonlinear, limit cycle oscillator of the form

$$\ddot{x} + (Vx^2 + R\dot{x}^2 - a)\dot{x} + kx = 0 \qquad (2)$$

to model these data, the covariation between frequency and amplitude is mimicked by changing only a single parameter, k, the linear restoring force (stiffness) of the oscillator (see Figure 2 for single wrist data, and Figure 3 for examples of observed and simulated movements in the time domain and on the phase plane). Note that this dynamic structure is actually a combination of the classic van der Pol and Rayleigh oscillators, which are also shown in Figure 2. These differ in the form of the nonlinear damping term. For the van der Pol oscillator,

$$\ddot{x} + (Vx^2 - a)\dot{x} + kx = 0 \qquad (3)$$

amplitude remains constant across changes in oscillator frequency; that is, the frequency-amplitude function (see Figure 2) has a finite y- (amplitude-) intercept, but the slope is everywhere zero. For the Rayleigh oscillator,

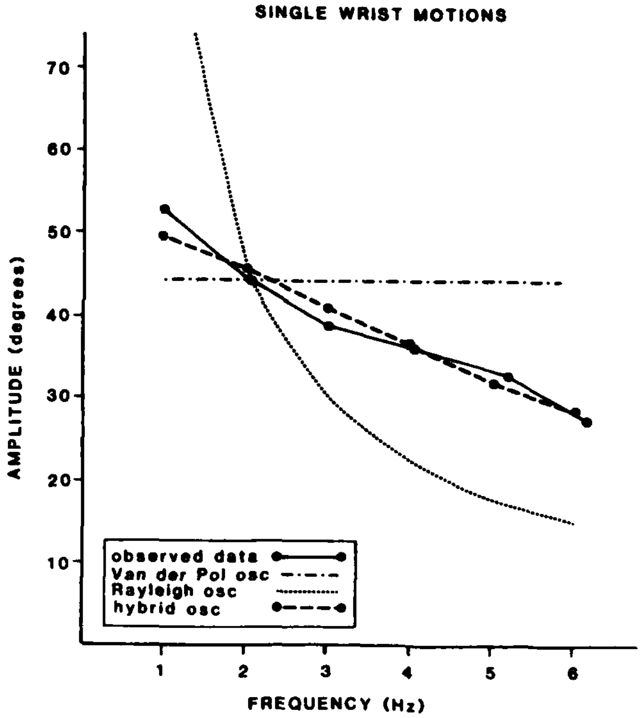


Figure 2. Amplitude-frequency relationship for single hand movements around the wrist joint, and three oscillator models (Kay, Kelso, Saltzman, & Schöner, submitted).

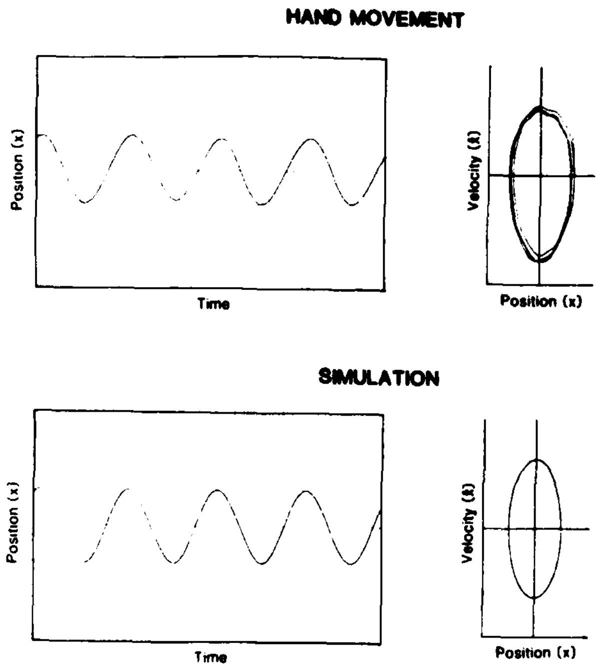


Figure 3. Time series and phase plane plots of a real hand movement (top) and a hybrid oscillator (bottom), both operating at 3 Hz.

$$\ddot{x} + (B\dot{x}^2 - a)\dot{x} + kx = 0 \qquad (4)$$

on the other hand, amplitude is inversely proportional to frequency, that is, the slope is everywhere negative but the y-intercept is infinitely large. (Infinite movement amplitude at zero movement frequency seems very unrealistic, both from intuition and our data.) The hybrid dynamics (equation (2) above) map onto the real data rather well both in terms of slope and intercept. When two such hybrid oscillators are coupled together (via terms proportional to the other oscillator's position and velocity, which is thus a linear coupling structure), once more a variation in system stiffness produces space-time behavior mimicking that observed for the two modes, mirror (in phase) shown in Figure 4, and parallel (anti-phase) shown in Figure 5.

Although the physiological underpinnings of the nonlinear parameters (or indeed system stiffness, assumed to be linear here) are opaque at the moment, these models allow us to make a simple, but we think important, point: Namely, that what we illustrate here is how a rather simple dynamical control structure, requiring variations in only one system parameter, can describe the spatiotemporal behavior of the limbs singly and together. It should not be lost on the reader that, regardless of its physiological origins, the nonlinearity is crucial to guarantee the particular frequency-amplitude relationship observed.

5.3 Generative Properties--Bifurcations

Fixed point and periodic attractors, as illustrated above, generate some of the behavioral characteristics observed in discrete and rhythmical movements, respectively. A nontrivial correspondence between model and reality is the feature of the stable behavior in spite of perturbations and small changes in parameters. Thus the shape of a limit cycle may change a bit or the time needed to complete a cycle may exhibit small variations as a parameter is varied. In such cases, the attractor can be said to change smoothly without altering its topological form. However, the topology of an attractor may change abruptly--a distinct change to a new form may occur--when a key parameter crosses a bifurcation point. At the bifurcation point (after the Latin, to branch), the system's behavior is ill-defined; it may show the old behavior or the new one. For example, Figure 1c shows the bifurcation diagram of the much-studied Hopf bifurcation (for many illustrations see Cvitanovic, 1984). On the phase plane (see Section 4 above), the system exhibits only a point stability at first, but upon changing the key parameter, μ, of the system past a certain value, a limit cycle trajectory ensues, as well as an unstable fixed-point. In Figure 1c, the straight line represents an equilibrium or steady state solution, for values of $\mu < \mu_c$. At the critical point μ_c, the system loses its prior stability--a steady state becomes oscillatory, as illustrated by the circle. A similar bifurcation--called the Pitchfork bifurcation--is shown in Figure 1d (see e.g., Haken, 1983). Here again a stable fixed point loses its stability and gives rise to a stable periodic orbit as the parameter is changed.[2]

Similar phenomena abound in nature, including biological motion, from the transitions in phase observed in simple materials (e.g., from solid to liquid to gas) to the transitions in gait patterns observed in horses (walk to trot to gallop, see Hoyt & Taylor, 1981) to transitions in human posture (see Nashner & McCollum, 1985; and, for a bifurcation interpretation, Saltzman & Kelso, 1985). Parametrically scaled bimanual movements have been shown to exhibit bifurcation (Kelso, 1981, 1984). Thus, starting in an antiphase modal

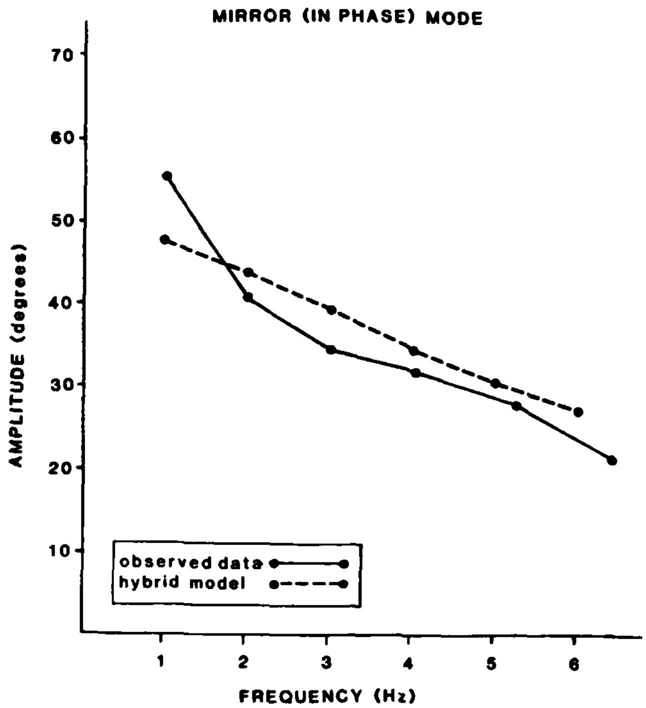


Figure 4. Amplitude-frequency relationship for in-phase (two-handed) movements, and the coupled hybrid oscillator model (from Kay et al., submitted).

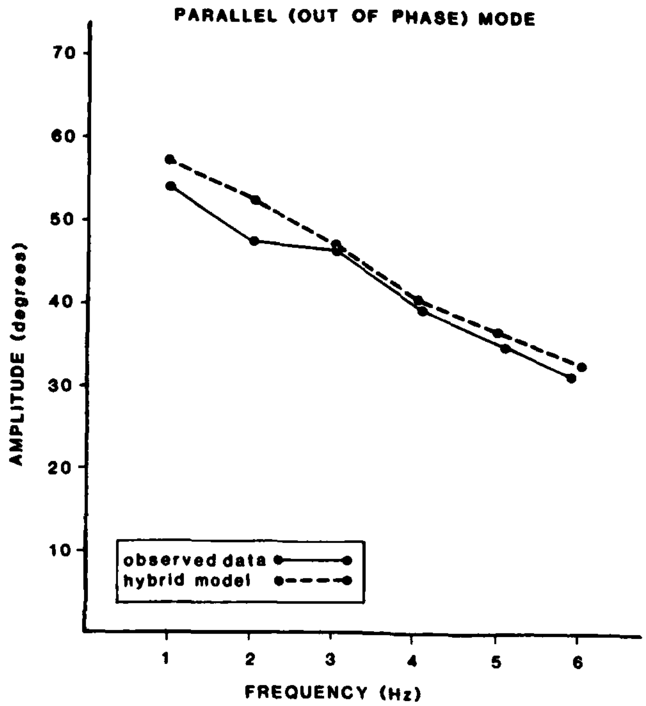


Figure 5. Amplitude-frequency relationship for anti-phase (two-handed) movements, and the coupled hybrid oscillator model (from Kay et al., submitted).

pattern (i.e., right flexion [extension] accompanied by left extension [flexion]), subjects in Kelso's studies voluntarily increased the cycling frequency of the two hands in a continuous manner. As frequency increased, the antiphase mode became less stable, as indicated by an increase in phase variance between the hands. At a critical parameter value (which the data suggested to be a dimensionless function of each individual's preferred cycling rate) the system bifurcated, and a different, in-phase modal pattern emerged. Though not given a bifurcation interpretation, similar results have been obtained by Baldissera, Cavallari, and Civaschi (1982), Cohen (1971), and MacKenzie and Patla (1983).

The bifurcation diagram shown in Figure 6 reflects the basic results of the Kelso experiments. If the bimanual system is "prepared" in the antiphase

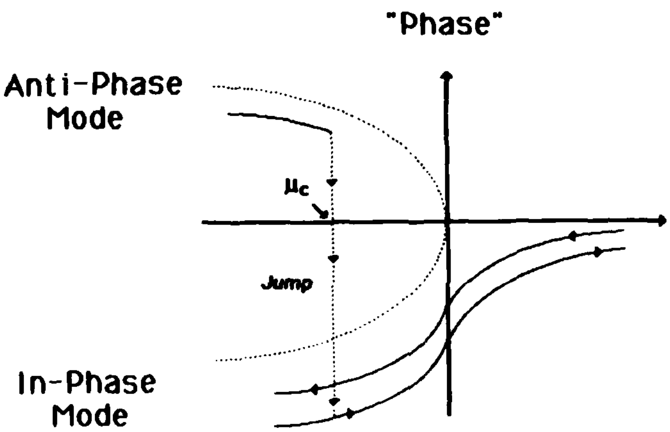


Figure 6. Bifurcation diagram of the bimanual phase transition: as the parameter μ is increased, the anti-phase mode becomes unstable (dashed lines), the in-phase mode stable. If μ is then decreased, behavior remains in the in-phase mode, i.e., the system stays on the same branch of the bifurcation picture.

mode (upper left quadrant), loss of stability occurs at the parameter value μ_c, i.e., when cycling frequency reaches a critical point, and a switch to the in-phase modal pattern occurs. The system then remains on the stable branch as μ is further increased (at least within limits). A further feature of the experiments shown in Figure 6, is that when cycling frequency is reduced, the system remains in the symmetric, in-phase mode, i.e., it exhibits the phenomenon of hysteresis. Using nonlinear oscillators similar to the one described in (2), and a <u>nonlinear</u> coupling[3] between them, Haken, Kelso, and

Bunz (1985) have explicitly modeled bimanual phase transition behaviors and generated novel, but testable predictions regarding their underpinnings.

In summary, we have illustrated here how it is possible for simple dynamical structures to generate a diversity of stable kinematic forms within a restricted region of their parameter spaces. In addition, we have shown how it is possible to explain the sometimes abrupt emergence of new kinematic forms when a critical bifurcation point is reached and the system enters a different region of parameter space. This analysis also hints at a kind of universal experimental strategy, viz., "tweak" system-sensitive parameters (externally or internally) to discover "new" spatiotemporal patterns. One is tempted to think that this is precisely what the emergence of skill is all about, and, parenthetically, what gifted teachers and coaches are all about as well. For it is they that often do the "tweaking" and it is they that have differentiated and become attuned to what some of the key parameters are (see Chapters 10-13 in Kelso, 1982b).

5.4 Inferring Dynamic Structure from Kinematic Analysis

A problem for investigators is that the dynamic parameters themselves are seldom, if ever, directly observed but can only be inferred from kinematic events. How can we go from kinematics to dynamics? By looking at key relationships (or relational invariants; see Kelso, 1981) among kinematic variables, one can gain valuable insights into the nature of the dynamics. For example, the mass-nonlinear spring system,

$$m\ddot{x} + kx + lx^3 = 0, \tag{5}$$

shows an invariant relationship between frequency and amplitude, depending on the sign of l, the nonlinear restoring force parameter. If l is positive, the spring force is termed "hard" since for larger amplitudes, the observed frequency is higher than for smaller amplitudes, and if negative, it is a "soft" spring with larger amplitude movements being slower than smaller amplitude ones (Jordan & Smith, 1977).

Kelso, Putnam, and Goodman (1983) applied the "soft"-spring model to their data on two-handed discrete movements of different amplitudes (see also Corcos, 1984; Marteniuk & MacKenzie, 1980; Marteniuk, MacKenzie, & Baba, 1984). The slight differences in movement time between simultaneously initiated short and long movements of the two limbs fall out, as it were, from a nonlinear model in which stiffness decreases with increasing distance from the equilibrium position. Thus movements of large amplitude will be slightly slower than those of short amplitude, because they have smaller average stiffnesses over the range of motion. Moreover, a prediction of this model--yet to be tested--is that the greater the amplitude differences between the two limbs the greater should be deviations from isochrony.

In the case of cyclical movements, the hybrid oscillator of Equation (2) displays the frequency-amplitude relationship observed in the Kay et al. (submitted) data (see Haken et al., 1985). The importance of nonlinearities is apparent here: autonomous oscillators (i.e., without explicit time-dependent forcing terms) with only linear springs and linear damping terms show no preferred relationship between frequency and amplitude. If, phenomenologically, there is some tight correlation between space and time, for example, then immediately nonlinear dynamics have to be invoked, the particular form of such a relationship giving insight into the particular

nature of such dynamics. In this sense, observed kinematic relationships between amplitude and frequency allow us to infer underlying dynamical control structures.

Another way of uncovering the dynamic control structure is to use kinematic relations evident in phase plane trajectories (see Section 4 above) to index dynamic parameters. For example, in a system of constant mass, the slope of the peak velocity-displacement relationship provides an estimate of system stiffness. A recent kinematic study by Kelso, V.-Bateson, Saltzman, and Kay (1985) of reiterant speech, (where a subject inserts a simple syllable /ba/ for real syllables in an utterance, performed at different rates) revealed a very systematic scaling relation between an articulatory gesture's peak velocity and its displacement. The finding that the relationship is linear throughout the movement range indicates that the stiffness is constant, supporting the notion that an invariant underlying dynamic is present. Further quantitative analysis of articulatory movement as a function of speaking rate and stress showed that both could be accounted for by a model with only two controllable parameters, system stiffness and equilibrium position. Preliminary modeling was consonant with this perspective. A major implication of the Kelso et al. (1985) studies (as well as much other evidence from unimanual and bimanual motor skills, some of which is discussed earlier) is that time _per se_ is not directly controlled; rather it is a consequence of the system's dynamic structure and parameterization.

Many other systems besides the lip-jaw complex exhibit a linear relationship between peak velocity and amplitude, for example, natural reaching movements (Jeannerod, 1984), drawing and handwriting (Lacquanti, Terzuolo, & Viviani, 1983; Viviani & McCollum, 1983), violin bowing (Nelson 1983), trombone playing (Wadman, Denier van der Gon, Geuze, & Mol, 1979), tongue movements (Ostry & Munhall, 1985), and eye movements (Bahill, Clark, & Stark, 1975). One can imagine that the structures involved all share the fundamental property of elasticity: any strains imposed upon them are met by linearly proportional forces, a force-displacement law that is precisely stated by the mass-spring dynamic. These examples show that a single dynamic structure can hold quite generally across a wide range of material structures sometimes involving multiple degrees of freedom and in many different kinds of action. Importantly, the data illustrate how kinematic relations can be used to infer (or as we prefer to say, to specify) dynamics.

5.5 Some Hard Problems for the Dynamical Approach

The above sections seem to promise a bright future for the dynamical approach to movement control. However, some problems stand in the way of success. First, given that we are looking at dynamical systems when we are observing organisms behaving, how can we be sure of the _uniqueness_ of the descriptions we apply? Many dynamical structures can give rise to similar kinematic consequences: for example, limit cycle-type behavior is exhibited both by nonlinear autonomous oscillators of the form of equation (2), and by the forced Duffing equation

$$m\ddot{x} + b\dot{x} + kx + lx^3 = f\cos(\omega t), \qquad (6)$$

which contains a time-dependent forcing term on the right hand side, rendering the equation nonautonomous and therefore very different in structure from the hybrid oscillator. Both of these oscillators settle down to an invariant limit cycle trajectory, and return to that cycle after perturbation.

Distinguishing these two options on the basis of actual behavior is problematical, but hope lies in the fact that other behavioral properties differ. In particular, the forced Duffing oscillator shows a jump in amplitude at a certain frequency, whereas the hybrid oscillator shows no such discontinuity in its frequency-amplitude curve.

Given the possibility of multiple dynamical descriptions, one of the investigators' tasks in the dynamical approach is to become familiar with the behavioral characteristics of various classes of dynamical systems and to obtain data addressing their similarities and differences. This is the approach we have taken in our work. The reader should beware, however, of the difficulties involved. Dynamics typically starts with a set of equations and evaluates their solutions under various conditions such as changes in parameters and initial conditions. Nonlinear dynamical systems, however, generally defy exact solutions and only approximate (via numerical methods) and/or qualitative solutions are possible. In movement science we are faced with an even more difficult problem: given a solution--a particular spatiotemporal event produced by an organism in an environment--what kinds of equations would produce this particular solution? This is where dynamical analogy (see Section 4 above) seems so crucial: an insight is needed into the similarity between the real event and something we know--such as a nonlinear oscillator. Then, when the latter is appropriately adapted, at least a qualitative model of the data becomes possible.

Another problem concerns the role of information in a dynamical system. In Section 5.1 above we argued that a functional grouping of muscles exhibits behavior qualitatively similar to a (nonlinear) mass-spring system. Such systems are intrinsically self-equilibrating in the sense that the end-point of the system or its "target" is achieved regardless of *initial conditions*. In such a model, the target is not achieved by means of conventional, closed-loop control, though targeting behavior can certainly be described by such a system. But sensory feedback, comparators, and reference levels have no role whatsoever in the dynamical systems considered here.

However, this is not to say that propriospecific information is unimportant--only to raise the question of how it is to be conceptualized and used within the present framework. As elaborated by Kelso, Holt, and Flatt (1980), standard views of peripheral mechanoreceptors are that they provide feedback about variables such as position, rate, and acceleration. Such feedback in a closed-loop system is referential to a structural entity, typically a setpoint that the system is trying to attain. Regulation and control are then effected by means of error detection and correction processes. There are good reasons to believe that this view has been greatly overvalued for biological systems. For example, although recognizing that setpoints can play a useful role in certain engineering applications, Cecchini, Melbin, and Noordergraaf (1981) state with reference to biological control that "there is no basis to conclude the existence of separate structural entities ... that define setpoints" and that setpoints are better considered "an arbitrary convenience" (p. 393; see also Kelso, 1981; Kugler et al., 1982; Yates, 1979).

As discussed in Section 3, we believe that a conception of information is required that is unique and specific to the state of the system's dynamics (Kelso, Holt, Kugler, & Turvey, 1980; Kugler et al., 1980). It is possible that such information is not given in terms of dimension-specific receptor codes but rather in geometrical terms, that is, in the form of the gradients

and equilibrium points in the system's potential energy function, which is an alternative representation of its dynamic structure (e.g., Hogan, 1980; Kugler et al., 1980). A task's potential energy function can be visualized as a surface with various hills and valleys, hills corresponding to less stable states, valleys to more stable states. Recently Fel'dman and Latash (1982) have presented a model emphasizing the intrinsic relationship between afferent and efferent signals in postural control that they feel "is in good correspondence with ideas [expressed here and elsewhere] about the dynamic nature of motor control and with the general concept that information in the nervous system reflects different forms of dynamic state and intrinsic metrics of control" (p. 188). This view of information as geometrically and/or topologically specified in the system's dynamic qualities is obviously novel (Thom, 1975) and has yet to be fully explored, but it offers an alternative to simplistic coding schemes in which receptor signals on a single dimension are fed back to a setpoint or a system comprised of multiple setpoints.

Interestingly, it appears that the dynamical approach is now being exploited in robotics research. In a recent conference entitled "Robotics Research: The Next Five Years and Beyond," Coleman (1985) reports that new methods for path planning are now being successfully implemented. Path planning has conventionally required that the robot possess a world model of its environment and a complex series of algorithms to compute the optimal path through (or around) a series of obstacles. Such methods require a prior representation of the entire work space, which often cannot be known completely in advance. Moreover, this kind of path planning is complicated from a computational point of view and does not produce good trajectories.

The new alternative--entirely consonant with the discussion above--is called the potential field approach and eliminates many of the problems of conventional methods. To guide the robot through a cluttered environment requires the specification of two sets of objects, goals and obstacles, which have potential fields associated with them (akin to a magnet's magnetic field). A goal, like a task in Saltzman and Kelso (1983b), is defined by an attractor (whose strength and direction are a function of its parameters), whereas the strength and direction of an obstacle are defined by an avoidance vector (or in dynamical language, a repellor). The sum of the attractor and avoidance vectors creates an acceleration vector for the robot to follow. Adaptive changes to the environment are also possible. Apparently, this method can be shown not only to reduce the computational complexity typical of path planning approaches, but also to improve considerably the quality of the resultant trajectories.

In addition, the view that information is available in the geometry of the system's dynamics also has been voiced by Boylls and Greene (1984) in their assessment of Bernstein's (1967) significance for the movement field today. With reference to impedance or endpoint control, they hypothesize that such theories

> ..will soon be recast in terms of potential functions (with endpoints identifiable as the extrema of such functions to be "sought," gradient fashion by the state of the skeletomotor system) [p. xxiii, emphasis ours]

Clearly, this view of propriospecific information is not anything like conventional notions of sensory feedback, and we can look forward to its elaboration in the near future. Moreover, a different image of

perceptual-motor learning is suggested--one in which the learner actively explores a task's potential energy function in order to discover its topology and identify its extrema. Learning (from the learner's perspective) is a problem of becoming sensitive to the information carried in the gradients and equilibrium points of potential surfaces (see Fowler & Turvey, 1978; Kugler, 1983).

A final problem considered here is that nonlinear dynamics classifies its attractors, by definition, in terms of families of trajectories and their asymptotic behavior. On the one hand this is a very powerful strategy, but on the other it begs the question of how a particular trajectory is elected. Once the dynamic parameters are set up for a task and the initial conditions defined, the dynamical approach provides a good account of the space-time behavior of the movement system. But how are the necessary conditions established? In the next section we look to the world of perception for insights into this issue.

6. Control of Action Dynamics Via Perception (Kinematic Specification)

In the above sections we have shown that dynamics can serve as a rich framework for theories of control, in that it affords low-dimensional control possibilities, and yet can generate a wide variety of behavior. We have also shown how the dynamics of movement control can be studied, via analysis of kinematic invariant relationships. We now come to the rather difficult problem raised in the previous section: how do the dynamic structures underlying action arise, and how are they modulated (i.e., how are their parameters set)? It has been argued (e.g., Runeson & Frykholm, 1982; Turvey, 1977; Turvey, Shaw, Reed, & Mace, 1981; Warren & Shaw, 1981) that perception provides the properties necessary to solve this problem for animals. However, perceptual events involve no forceful interactions: the events occurring in the flow of the optic field, for example, are purely kinematic in nature (Gibson, 1966, 1979; Runeson, 1977). Similar to the problem investigators have in determining the dynamics of action, organisms have the problem of perceiving the dynamic structure of events solely from the kinematic array. But, as illustrated in the following examples, critical properties of kinematic flow fields define information specific to the dynamical interactions of organism and environment (Runeson & Frykholm, 1982; Yates & Kugler, in press).

Consider the problem of driving a car up to an intersection. There are two ways to stop the car: 1) by forceful interaction, e.g., by hitting a nearby tree; or 2) by using the flow of optical texture in the visual field to determine when contact might occur and what to do to avoid contact (Yates & Kugler, in press, provide this example). Lee (1976) has identified the kinematic property of the optic flow field that specifies time-to-contact of an object approaching an observer at a constant relative velocity along the line of sight. The rate of magnification of the object relative to the point of observation is this significant optical property. After Lee (1976, 1980) we can designate the inverse of this variable as τ (tau), the time-to-contact itself. Tau's importance is that it is a directly available, non-derived property of the optical flow field itself. Its powerful role in the guidance of biologically significant activities has been demonstrated in numerous studies (see von Hofsten & Lee, 1985; Lee, 1980; Lee & Young, in press; Solomon et al., in press; Turvey & Kugler, 1984; for reviews). For example, the gannet is a large seabird that dives for its prey from considerable heights, at variable speeds, and in the face of changing wind conditions.

Since the gannet is accelerating under gravity, if its wings were not retracted appropriately, it would annihilate itself upon hitting the water's surface. However, the gannet has been shown to be remarkably sensitive to τ and, in fact, initiates wing retraction when τ reaches a certain critical value.

Relatedly, flies have been shown to begin to decelerate prior to contact with a surface at a critical value of τ (Wagner, 1982). In addition, Wagner shows that no other combination of kinematic variables (which might feasibly be picked up perceptually) is as effective as τ. Returning to our driving example, Lee (1976) has further demonstrated that τ and its rate of change $\dot{\tau}$ provide the necessary information to avoid collisions with an obstacle. Thus the value of $\dot{\tau}$ specifies whether braking is sufficiently hard: below $\dot{\tau} = -.5$, safety is assured. Above it, however, the applied decelerative forces are inadequate to avoid collision. From these examples, we see that τ and its rate of change are key parameters for the regulation of action. Not only do they provide continuous information for modulating activity, but they also effect bifurcations to different (and adaptive) modes of behavior.

Time-to-contact is not the only aspect of the optic field that has been found to regulate actor dynamics. Warren (1984) had short and tall subjects visually rate the "climbability" of sets of stairs of varying riser heights. He found that observers of widely different dimensions chose those stairs that optimally matched their body size. The measure of "sameness" in this case was intrinsic to the observer, i.e., the same ratio of riser height/leg length indexed climbability in both tall and short people. This ratio is an intrinsic metric akin to the time-to-contact variable τ in the above examples. According to Warren (1984), two competing factors may determine the fit between organism (climber) and environment (stairs) in this task. As the ratio of riser height to leg length increases, more energy must be expended to raise the subject's body mass a given vertical distance. On the other hand, as the ratio decreases more steps must be made to accomplish the same amount of work. These competing tendencies may serve to establish an optimum point of minimum metabolic demand for the organism-environment system. Warren found that subjects differed greatly in their oxygen consumption when climbing a series of moving, escalator-like stairmills (analogous to a treadmill) whose tread-to-riser height was varied. However, when the data were scaled to conform with the subjects' body dimensions, the oxygen consumption minimum occurred at precisely the same ratio that corresponded to their preferred perceptual judgments. In Warren's work we see a beautiful example of optical specification in body-scaled (intrinsic) terms, providing the observer with information about the fit between his or her dimensions and the stair (see also Warren & Kelso, 1985; Warren & Shaw, 1985, for reviews). In addition and importantly for the present discussion, Warren shows that by enlarging the frame of reference to include animal and environment, perceptual category boundaries (critical points)--separating climbable and nonclimbable stairs--are also predicted by his biomechanical model.

7. Common Principles Linking Dynamic Events in Perception and Action?

Drawing from many of the examples presented in Sections 5 and 6 we see some impressive parallels between the dynamics of movement control and the perception of dynamic events. Remember, the thrust of this paper as with much of the work referred to herein has been to identify (relatively abstract) functional organizations common to structurally very different subsystems. The equivalence between the behavior of a complex neuromuscular system and a

nonlinear oscillator, as discussed in Section 4, is abstract and functional, rather than concrete and structural. In the context of this paper such an approach seeks principles that apply not just to movement control or perception alone, but to the perception-action system as a whole. Could it be that perception and action--typically treated as independent domains of inquiry--are really coupled by virtue of sharing common (dynamical) principles? If so, what are they?

We saw above that the optic flowfield is literally a global morphology (a velocity vector field) or form that uniquely informs (in the sense of Varela, 1979, and Section 2 above) the organism of the many ways it can adjust to its environment. Real or artificially-induced global optical changes can be shown to produce lawfully related perceptual experiences. Similarly, in Section 5 we saw how the forms of motion, given in phase portraits, allow the scientist to uncover an underlying dynamical control structure. In both cases it is the form of the kinematics that informs--in the sense of a lawful mapping--dynamical states of affairs. We say, after Runeson, that kinematics specify dynamics.

We saw, in Section 5, that a criterion for the stability of an attractor is that it exhibit smoothness in the face of parameter changes and perturbations. But we also saw that when a parameter crosses a critical threshold, bifurcation occurs--there is a switch from one type of behavior to another. Literally, a behavioral phase transition occurs. Both perception and action subsystems share the features of stability on the one hand and criticality on the other. Which behavior is observed depends on which regions of the parameter space the system occupies. From Warren's and others' work we see that stable and critical behavior arise not just in the perception and action subsystems individually, but arise from the dynamics of the animal-environment system as a unit.

The individual analyses of production and perception show how enormously detailed microscopic descriptions are, in each case, reduced to low-dimensional, macroscopic descriptions. In Lee, Lishman, and Thompson's (1982) analysis of skilled long jumping we see a conflation of macroscopic parameters. Only one macroscopic optical property appears to be pertinent to the jumper's adjustment to the upcoming board, the time-to-contact, τ. And only one macroscopic movement parameter appears to reflect the jumper's motoric adjustments, the impulse generated during the stance phase of the gait cycle. Thus a highly complex control problem reduces to a coupling between just two macroscopic parameters (see also Fitch, Tuller, & Turvey, 1982; Solomon et al., in press). Whether other tasks are amenable to a similar kind of analysis is open to question. Kelso et al. (1985) suggest that the stiffness changes they observe between stressed and unstressed speech gestures may specify listener's perception of stress, an hypothesis that can be tested directly by articulatory synthesis (see, e.g., Browman, Goldstein, Kelso, Rubin, & Saltzman, 1984). Similarly, the phasing structure of articulatory movements may map directly onto listener's perception of speaking rate (Kelso, 1985).

Are the various parallels mentioned here between perception and production just that, parallels, or is there a deeper dynamical structure linking them together? A quote from Feynman's (1967) classic, The Character of Physical Law, may leave the reader with an impression of our position:

> This kind of game of roughly guessing at family relationships...is illustrative of the kind of preliminary sparring which one does with

nature before really discovering some deep and fundamental law (p. 155).

Action and perception have evolved together. Just because we analyze them separately is no reason to divorce them from each other, or not to search for the lawful basis of their linkage.

References

Abraham, R. H., & Shaw, C. D. (1982). Dynamics--The geometry of behavior. Santa Cruz, CA: Aerial Press.

Anderson, P. W. (1972). More is different. Science, 177, 393-396.

Andronov, A., & Chaiken, C. E. (1949). Theory of oscillations. Princeton, NJ: Princeton University Press.

Ashby, W. R. (1956). An introduction to cybernetics. London: Metheun & Co., Ltd.

Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence: A tool for studying human eye movements. Mathematical Biosciences, 24, 191-204.

Baldissera, F., Cavallari, P., & Civaschi, P. (1982). Preferential coupling between voluntary movements of ipsilateral limbs. Neuroscience Letters, 34, 95-100.

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon Press.

Bizzi, E., Chapple, W., & Hogan, N. (1982). Mechanical properties of muscles: Implications for motor control. Trends in Neurosciences, 5, 395-398.

Bizzi, E., Polit, A., & Morasso, P. (1976). Mechanisms underlying achievement of final head position. Journal of Neurophysiology, 39, 435-444.

Boylls, C. C., & Greene, P. H. (1984). Introduction: Bernstein's significance today. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed (pp. xix-xxix). Amsterdam: North-Holland.

Browman, C. P., Goldstein, L., Kelso, J. A. S., Rubin, P., & Saltzman, E. (1984). Articulatory synthesis from underlying dynamics. Journal of the Acoustical Society of America, 75, S22-S23. (Abstract)

Cecchini, A. B. P., Melbin, J., & Noordergraaf, A. (1981). Setpoint: Is it a distinct structural entity in biology control? Journal of Theoretical Biology, 93, 387-394.

Cohen, A. H., Holmes, P. J., & Rand, R. H. (1982). The nature of the coupling between segmental oscillators of the lamprey spinal generator for locomotion: A mathematical model. Journal of Mathematical Biology, 13, 345-369.

Cohen, L. (1971). Synchronous bimanual movements performed by homologous and non-homologous muscles. Perceptual and Motor Skills, 32, 639-644.

Coleman, A. (1985). Robotics research: The next five years and beyond. Robotics Age, 7(2), 14-19 (February).

Cooke, J. D. (1980). The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland.

Corcos, D. M. (1984). Two-handed movement control. Research Quarterly For Exercise and Sport, 55, 117-122.

Cvitanovic, P. (Ed.) (1984). Universality in chaos. Bristol, U.K.: Adam Hilger.

Fel'dman, A. G. (1966). Functional tuning of the nervous system with control of movement or maintenance of a steady posture. III. Mechanographic

analysis of execution by man of the simplest motor tasks. *Biophysics, 11*, 766-775.

Fel'dman, A. G., & Latash, M. L. (1982). Interaction of efferent and afferent signals underlying joint position sense: Empirical and theoretical approaches. *Journal of Motor Behavior, 14*, 174-193.

Feynman, R. (1967). *The character of physical law.* Cambridge, MA: MIT Press.

Fitch, H. L., Tuller, B., & Turvey, M. T. (1982). The Bernstein perspective: III. Tuning of coordinative structures with special reference to perception. In J. A. S. Kelso (Ed.), *Human motor behavior: An introduction*. Hillsdale, NJ: Erlbaum.

Fitch, H. L., & Turvey, M. T. (1978). On the control of activity: Some remarks from an ecological point of view. In D. Landers & R. Christina (Ed.), *Psychology of motor behavior and sports*. Urbana, IL: Human Kinetics.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381-391.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press.

Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press.

Freund, H. -J. (1983). Motor unit and muscle activity in voluntary motor control. *Physiological Reviews, 63*, 387-436.

Gibson, J. J. (1961). Ecological optics. *Vision Research, 1*, 253-262.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton-Mifflin.

Greene, P. H. (1982). Why is it easy to control your arms? *Journal of Motor Behavior, 14*, 260-286.

Guckenheimer, J., & Holmes, P. (1983). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. New York: Springer Verlag.

Haken, H. (1975). Cooperative phenomena in systems far from thermal equilibrium and in nonphysical systems. *Review of Modern Physics, 47*, 67-121.

Haken, H. (1977). *Synergetics: An introduction*. Heidelberg: Springer Verlag.

Haken, H. (1983). *Advanced synergetics*. Heidelberg: Spring-Verlag.

Haken, H., Kelso, J. A. S. & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics, 51*, 347-356.

von Hofsten, C., & Lee, D. N. (1985). Dialogue on perception and action. In W. H. Warren & R. E. Shaw (Eds.), *Persistence and change: Proceedings of the First International Conference on Event Perception.* Hillsdale, NJ: Erlbaum.

Hogan, N. (1980). Mechanical impedance control in assistive devices and manipulators. In *Proceedings of the Joint Automatic Control Conferences*, San Francisco, Vol. 1, TA-108.

Hollerbach, J. M. (1982). Computers, brains, and the control of movement. *Trends in Neurosciences, 5*, 189-192.

Houk, J. C., & Rymer, W. (1981). Neural control of muscle length and tension. In V. B. Brooks (Ed.), *Handbook of physiology, Section 1*

(Vol. II, Motor control, Part 1, pp. 257-323). Bethesda, MD: American Physiology Society.

Hoyt, D. F., & Taylor, C. R. (1981). Gait and the energetics of locomotion in horses. Nature, 292, 239-240.

Jeannerod, M. (1984). The timing of natural prehension movements. Journal of Motor ehavior, 16, 235-254.

Jordan, D. W., & Smith, P. (1977). Nonlinear ordinary differential equations. Oxford: Clarendon Press.

Kay, B., Kelso, J. A. S., Saltzman, E. L., & Schöner, G. (submitted). The space-time behavior of single and bimanual rhythmical movements: Data and model. Submitted to Journal of Experimental Psychology: Human Perception and Performance.

Keele, S. W. (1981). Behavioral analysis of movement. In V. B. Brooks (Ed.), Handbook of physiology Section 1 (Volume II, Motor control, Part 2, pp. 1391-1414). Baltimore, MD: American Physiological Society.

Kelso, J. A. S. (1977). Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 3, 529-543.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (1982a). Overview of skilled performance: Coming to grips with the jargon. In J. A. S. Kelso (Ed.), Human motor behavior: An introduction. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (Ed.). (1982b). Human motor behavior: An introduction. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative, 246, R1000-R1004.

Kelso, J. A. S. (1985, February). A note on coarticulation and related issues. Paper presented at Conference on Units of speech production and perception: Origins and functions of their coarticulation. New Haven, CT: Haskins Laboratories.

Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 43, 1183-1196.

Kelso, J. A. S., Holt, K. G., & Flatt, A. E. (1980). The role of proprioception in the perception and control of human movement: Toward a theoretical reassessment. Perception & Psychophysics, 28, 45-52.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 49-70). New York: North-Holland.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of nonlinear limit cycle oscillatory processes: Theory and data. Journal of Motor Behavior, 13, 226-261.

Kelso, J. A. S., Putnam, C. A., & Goodman, D. (1983). On the space-time structure of human interlimb coordination. Quarterly Journal of Experimental Psychology, 35A, 347-376.

Kelso, J. A. S., & Saltzman, E. L. (1982). Motor control: Which themes do we orchestrate? The Behavioral and Brain Sciences, 5, 554-557.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., & Tuller, B. (1984a). Converging sources of evidence for common dynamical principles in speech and limb coordination. American Journal of Physiology: Regulatory, Integrative, and Comparative, 246,

R928-935.

Kelso, J. A. S., & Tuller, B. (1984b). A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience (pp. 321-356). New York: Plenum.

Kelso, J. A. S., V.-Bateson, E., Saltzman, E. L., & Kay, B. (1985). A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling. Journal of the Acoustical Society of America, 77, 266-280.

Kugler, P. N. (1983). A morphological view of information for the self-assembly of rhythmic movement: A study in the similitude of natural law. Unpublished doctoral dissertation, University of Connecticut.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 3-47). New York: North-Holland.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1982). On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), The development of movement control and coordination (pp. 5-78). Chichester: John Wiley.

Kugler, P. N., & Turvey, M. T. (in press). Information, natural law, and the self-assembly of rhythmic movement: Theoretical and experimental investigations. Hillsdale, NJ: Erlbaum.

Lacquanti, F., Terzuolo, C., & Viviani, P. (1983). The law relating the kinematic and figural aspects of drawing movements. Acta Psychologica, 54, 115-130.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Lee, D. N., Lishman, J. R., & Thomson, J. A. (1982). Regulation of gait in long jumping. Journal of Experimental Psychology: Human Perception and Performance, 8, 448-459.

Lee, D. N., & Young, D. S. (in press). Gearing action to the environment. Experimental Brain Research Supplement.

MacKenzie, C. L., & Patla, A. E. (1983). Breakdown in rapid bimanual finger tapping as a function of orientation and phasing. Society for Neuroscience. (Abstract).

Marteniuk, R. G., & MacKenzie, C. L. (1980). A preliminary theory of two-handed movement control. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Marteniuk, R. G., MacKenzie, C. L., & Baba, D. M. (1984). Bimanual movement control: Information processing and interaction effects. Quarterly Journal of Experimental Psychology, 36A, 335-365.

Maxwell, J. C. (1952). Matter and motion. New York: Dover Press. (Original work published 1877)

Miller, S., & Scott, P. D. (1977). The spinal locomotor generator. Experimental Brain Research, 30, 387-403.

Minorsky, N. (1962). Nonlinear oscillations. Princeton, NJ: Van Nostrand.

Morasso, P. (1981). Spatial control of arm movements. Experimental Brain Research, 42, 223-227.

Morowitz, H. J. (1968). Energy flow in biology; Biological organization as a problem in thermal physics. New York: Academic Press.

Nashner, L. M., & McCollum, G. (1985). The organization of human postural movements: A formal basis and experimental synthesis. The Behavioral and Brain Sciences, 8, 135-172.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. Biological Cybernetics, 46, 135-147.
Ostry, D. J., & Munhall, K. (1985). Control of rate and duration in speech. Journal of the Acoustical Society of America, 77, 640-648.
Patla, A. E., Calvert, T. W., & Stein, R. B. (in press). A model of a pattern generator for locomotion in mammals. American Journal of Physiology.
Pattee, H. H. (1972). Laws and constraints, symbols and language. In C. H. Waddington (Ed.), Towards a theoretical biology. Chicago: Aldine.
Pattee, H. H. (1973). Physical problems of the origin of natural controls. In A. Locker (Ed.), Biogenesis, evolution, homeostasis (pp. 41-49). Heidelberg: Springer-Verlag.
Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. International Journal of General Systems, 3, 259-266.
Prigogine, I., & Stengers, I. (1984). Order out of chaos: Man's new dialogue with Nature. New York: Bantam.
Rasband, P. (1983). Dynamics. New York: John Wiley & Sons.
Reed, E. S. (1981). Indirect action. Unpublished manuscript, University of Minnesota, Center for Research in Human Learning.
Reed, E. S. (1982). An outline of a theory of action systems. Journal of Motor Behavior, 14, 98-134.
Rosen, R. (1970). Dynamical system theory in biology: Vol. 1. Stability theory and its applications. New York: Wiley.
Rosen, R. (1978). Fundamentals of measurement and representation of natural systems. New York: North-Holland.
Runeson, S. (1977). On visual perception of dynamic events. Unpublished doctoral dissertation, University of Uppsala.
Runeson, S., & Frykholm, G. (1982). Kinematic specification of dynamics as an informational basis for person and action perception: Expectations, gender recognition, and deceptive intention. Uppsala Psychological Reports, #324, Dept. of Psychology, Univ. of Uppsala, Sweden.
Saltzman, E. L., & Kelso, J. A. S. (1983a). Toward a dynamical account of motor memory and control. In R. Magill (Ed.), Memory and control of action (pp. 17-38). Amsterdam: North-Holland.
Saltzman, E., & Kelso, J. A. S. (1983b). Skilled actions: A task dynamic approach. Haskins Laboratories Status Report on Speech Research, SR-76, 3-50.
Saltzman, E., & Kelso, J. A. S. (1985). Stabilities, instabilities, and modes. The Behavioral and Brain Sciences, 8, 161-163.
Schmidt, R. A. (1982). Motor control and learning: A behavioral emphasis. Champaign, IL: Human Kinetics.
Schmidt, R. A., & McGown, C. (1980). Terminal accuracy of unexpectedly loaded rapid movements: Evidence for a mass-spring mechanism in programming. Journal of Motor Behavior, 12, 149-161.
Shaw, R. (1981). Modeling chaotic systems. In H. Haken (Ed.), Order and chaos in nature (pp. 218-231). New York: Springer-Verlag.
Solomon, J., Carello, C., & Turvey, M. T. (in press). Flow fields: Optical support for skilled activities. In W. F. Straub & J. Williams (Eds.), Cognitive sports psychology. Lansing, NY: Sport Science Associates.
Stein, R. B. (1982). What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, 5, 535-577.
Thom, R. (1975). In D. H. Fowler (Trans.), Structural stability and morphogenesis. Reading, MA: Benjamin, Inc.
Tomovic, R. (1978). Some central conditions for self-organization--what the control theorist can learn from biology. American Journal of Physiology: Regulatory, Integrative, and Comparative, 235, R205-R209.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

Turvey, M. T., & Kugler, P. N. (1984). Information and the ecological approach to perception and action. In H. T. A. Whiting (Ed.), Human motor actions: Bernstein reassessed. Amsterdam: North Holland.

Turvey, M. T., Shaw, R. E., Reed, E. S., & Mace, W. M. (1981). Ecological laws of perceiving and acting: In reply to Fodor and Pylyshyn (1981). Cognition, 9, 237-304.

Varela, F. J. (1979). Principles of biological autonomy. New York: North-Holland.

Viviani, P., & McCollum, G. (1983). The relation between linear extent and velocity in drawing movements. Neuroscience, 10, 211-218.

Wadman, W. J., Denier van der Gon, J. J., Geuze, R. H., & Mol, C. R. (1979). Control of fast goal-directed arm movements. Journal of Human Movement Studies, 5, 3-17.

Wagner, H. (1982). Flow-field variables trigger landing in flies. Nature, 297, 147-148.

Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. Journal of Experimental Psychology: Human Perception and Performance, 10, 683-703.

Warren, W. H., & Kelso, J. A. S. (1985). Report of the work group on perception and action. In W. H. Warren & R. E. Shaw (Eds.), Persistence and change: Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum.

Warren, W. H., & Shaw, R. E. (1981). Psychophysics and ecometrics. The Behavioral and Brain Sciences, 4, 209-210.

Warren, W. H., & Shaw, R. E. (Eds.). (1985). Persistence and change: Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum.

Whiting, H. T. A. (1984). Human motor actions: Bernstein reassessed. Amsterdam: North-Holland.

Willis, J. B. (1980). On the interaction between spinal locomotor generators in quadrupeds. Brain Research Reviews, 2, 171-204.

Yates, F. .E. (1979). Physical biology: A basis for modeling living systems. Journal of Cybernetics and Informations Science, 2, 57-70.

Yates, F. E., & Kugler, P. N. (in press) Signs, singularities, and significance: A physical model for semiotics. Semiotica.

Yates, F. E., Marsh, D. J., & Iberall, A. S. (1972). Integration of the whole organism: A foundation for a theoretical biology. In J. A. Behnke (Ed.), Challenging biological problems: Directions towards their solution (pp. 110-132). New York: Oxford University Press.

Footnotes

[1]Dynamical organizations can be used to categorize movement tasks into distinct topological forms. Topology is the branch of mathematics that categorizes, for example, geometrical shapes, on the basis of the loosest possible criterion: continuity of form. A circle, ellipse, square, and any simple closed curve in the plane are topologically indistinguishable, whereas a line and a circle fall into separate topological categories (although they are all one-dimensional curves). To transform a circle into a line requires breaking the circle, i.e., a change in the continuity of the circle. Applied to movement, the kinematics of tasks may be treated as shapes that can be put into topological classes, or topologies. Plotting position versus velocity on the phase plane, one can see that discrete movements to a target exhibit

asymptotic behavior to a point topology (hence are characterized as a point attractor), while repetitive movements are similar to circles and ellipses and form a periodic attractor topology. Other kinds of movements may require the definition of other topologies, e.g., a chaotic attractor (see Shaw, 1981) for physiological tremor. Different dynamical organizations can thus generate different movement topologies.

[2]Mathematically the difference between a Hopf and Pitchfork bifurcations rests in whether a pair of eigenvalues or a single eigenvalue, respectively, crosses the imaginary axis when the parameter passes through a critical value (see Eckmann, 1981), i.e., whether the bifurcation occurs, at a fundamental level, in a space of two dimensions or one.

[3]Terms proportional to the product of the position squared and velocity of the other oscillator, similar to a van der Pol damping structure were used. Current work is underway that tries to account for the previously mentioned frequency-amplitude data in terms of this nonlinear coupling structure. If successful, a single model would then describe both the stable and transition behavior.

PUBLICATIONS

APPENDIX

PUBLICATIONS

Alfonso, P. J., Watson, B. C., Baer, T., Sawashima, M., Hirose, H., Kiritani, S., Niimi, S., & Itoh, K. (1985). The organization of supralaryngeal articulation in stutterers' fluent speech production: A preliminary report. In Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (Vol. 19, pp. 191-200).

Beddor, P. S., Krakow, R. A., & Goldstein, L. M. (in press). Perceptual constraints and phonological change: A study of nasal vowel height. Phonology Yearbook, 3.

Borden, G. J., Baer, T., & Kenney, M. K. (1985). Onset of voicing in stuttered and fluent utterances. Journal of Speech and Hearing Research, 28, 363-372.

Brady, S. (in press). Short-term memory, phonological processing and reading ability. Annals of Dyslexia.

Cooper, A. M., Whalen, D. H., & Fowler, C. A. (in press). P-centers are unaffected by phonetic categorization. Perception & Psychophysics.

Hanson, V. L., & Wilkenfeld, D. (1985). Morphophonology and lexical organization in deaf readers. Language and Speech, 28, 269-280.

Krakow, R. A., & Hanson, V. L. (1985). Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print. Memory & Cognition, 13, 265-272.

Munhall, K. G. (in press). Motor equivalence and goal descriptors. The Behavioral and Brain Sciences.

Munhall, K. G., & Kelso, J. A. S. (1985). The role of similarity analysis in understanding movement. Journal of Motor Behavior, 17, 493-498.

Ostry, D. J., Cooke, J. D., & Munhall, K. G. (in press). Kinematic form of limb and speech movements. In G. N. Gantchev, B. Dimitrov, & P. Gatev (Eds.), Motor control. New York: Plenum Press.

Repp, B. H. (1985). Perceptual coherence of speech: Stability of silence-cued stop consonants. Journal of Experimental Psychology: Human Perception and Performance, 11, 799-813.

Repp, B. H. (in press). Some observations on the development of anticipatory coarticulation. Journal of the Acoustical Society of America.

Repp, B. H. (in press). Perception of the [m]-[n] distinction in CV syllables. Journal of the Acoustical Society of America,

Repp, B. H. (in press). Can linguistic boundaries change the effectiveness of silence as a phonetic cue? Journal of Phonetics.

Richards, J. T., & Hanson, V. L. (1985). Visual and production similarity of the handshapes of the American manual alphabet. Perception & Psychophysics, 38, 311-319.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. Journal of Memory and Language, 25, 123-135.

APPENDIX

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	AD A129713	ED 229-816
SR-74/75	April - September 1983	AD A136416	ED 236-753
SR-76	October - December 1983	AD A140176	ED 241-973
SR-77/78	January - June 1984	AD A145585	ED 247-626
SR-79/80	July - December 1984	AD A151035	ED 252-907
SR-81	January - March 1985	AD A156294	ED 257-159
SR-82/83	April - September 1985	AD A165084	**

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER SR-84, 1985	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Haskins Laboratories Status Report on Speech Research	5. TYPE OF REPORT & PERIOD COVERED Continuation Report Oct. 1-Dec. 31, 1985
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Staff of Haskins Laboratories: Alvin M. Liberman, President	8. CONTRACT OR GRANT NUMBER(s) HD-01994 NS13870 N01-HD-5-2910 NS13617 RR-05596 NS18010 BNS-8111470 N00014-83-K-0083
9. PERFORMING ORGANIZATION NAME AND ADDRESS Haskins Laboratories 270 Crown Street New Haven, CT 06511-6695	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS National Institutes of Health National Science Foundation Office of Naval Research	12. REPORT DATE January, 1986
	13. NUMBER OF PAGES 282
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) As above	15. SECURITY CLASS. (of this Report) Unclassified
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

UNLIMITED: Contains no information not freely available to the general public. It is distributed primarily for library use.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

As above

18. SUPPLEMENTARY NOTES

N/A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Speech Perception: phonological acquisition; phonological change; review
invariance; aspiration; English; voiceless consonants;
phonetic boundaries; phonetic cues; nasal distinctions;
cues;
melody-text; memory; song; integration;
language behavior,

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report (1 April-30 September) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

- Task dynamic coordination of the speech articulators: A preliminary model
- Some observations on the development of anticipatory coarticulation
- The role of production variability in normal and deviant developing speech
- Can linguistic boundaries change the effectiveness of silence as a phonetic cue?
- Perception of the [m]-[n] distinction in CV syllables
- On the nature of melody-text integration in memory for songs
- Some developments in research on language behavior

DD FORM 1473 1 Jan 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. Key Words (Continued)

Speech Articulation:
variability, deviation, speech development anticipation, coarticulation

Motor Control:
action coupling, perception, analysis, cooperative phenomena, motion task dynamics, coordination, articulation, model

Reading:
orthographic structure, acquisition, deafness, speech disability, object-naming, phonological deficiency learning, writing

20. Abstract (Continued)

-The pursuit of invariance in speech signals
-How is the aspiration of English /p,t,k/ "predictable"?
-Development phonology: Is the child father to the man?
-Phonology and the problems of learning to read and write
-Phonological deficiencies in children with reading disability: Evidence from an object-naming task
-Access to spoken language and the acquisition of orthographic structure: Evidence from deaf readers
-Cooperative phenomena in biological motion
-Naturalizing the context for interpreting SMA function
-Information and control: A macroscopic analysis of perception-action coupling

END

DTIC

8-86